Math Camp Notes
August 2025

Carlos Brunet Martins-Filho
Department of Economics
University of Colorado, Boulder
256 UCB
Boulder, CO 80309-0256
USA
email: carlos.martins@colorado.edu

1 Sets and basic operations

A set is a collection of arbitrary objects. These objects are called the elements of the set. A set is described by listing its elements or by enunciating the properties its elements must satisfy. It is common practice to represent a set by including its elements, or the properties satisfied by its elements, within curly brackets. For example, the set of all outcomes from rolling a 6-sided die, where the sides are numbered from 1 to 6, can be represented by

$$A = \{1, 2, 3, 4, 5, 6\} \text{ or by } A = \{x : x \text{ is a natural number and } 1 \leq x \leq 6\}.$$

In the second case, we read "*A is the set with typical element x such that x is a natural number and* $1 \leq x \leq 6$." Sets are commonly denoted by uppercase roman letters, such as A in this example.

In mathematics, some of the most fundamental sets contain numbers. The set of natural numbers is denoted by $\mathbb{N}$ with $\mathbb{N} = \{1, 2, \cdots\}$, the set of non-negative integers by $\mathbb{N}_0 = \{0, 1, 2, \cdots\}$, the set of integers by $\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$ and the set of rational numbers by $\mathbb{Q} = \{x : x = a/b \text{ where } b \neq 0 \text{ and } a, b \in \mathbb{Z}\}$. The set of real numbers, denoted by $\mathbb{R}$, is the set of rational numbers together with the set of irrational numbers, i.e., the numbers that are not rational.[1] Another important set contains the complex numbers and is denoted by $\mathbb{C}$. The set is formed by all real numbers together with what are known as "imaginary" numbers.[2]

If x is an element of a set A we write $x \in A$, and when it is not, we write $x \notin A$. If A and B are sets such that $x \in A \implies x \in B$, we write $A \subseteq B$ and say that A is a subset of B. If $A \subseteq B$ and $B \subseteq A$, we write $A = B$ and say that the two sets are equal, otherwise we write

[1] Showing that a particular number is irrational can be difficult, but the existence of irrational numbers is easy to establish. For example, $\sqrt{2}$ is an irrational number. To see this, suppose that it is not, such that we can write $\sqrt{2} = a/b$ where a and b have no common prime factors. Then $a^2 = 2b^2$ and a^2 is a multiple of 2. But since 2 is a prime number it divides a, so that there exists an integer c such that $a = 2c$ and $4c^2 = 2b^2$, which implies $b^2 = 2c^2$. Hence, b is a multiple of 2 which contradicts the assumption that a and b have no common factors. Thus, $\sqrt{2}$ cannot be a rational number.

[2] Complex numbers will be formally introduced and discussed later in these notes.

$A \neq B$. If $A \subseteq B$ and $A \neq B$ we say that A is a proper subset of B and when this needs to be emphasized we write $A \subset B$. The set that contains no elements is called the empty set and is denoted by $\emptyset$. The empty set is a subset of all sets, i.e., $\emptyset \subset A$ for any nonempty A. This is true because if the empty set is not a subset of A, then there exists an element of $\emptyset$ that is not in A. But this is impossible since $\emptyset$ has no elements. The set containing all subsets of a set A is called the power set of A and is denoted by 2^A or $\mathcal{P}(A)$. Sets whose elements are sets are often called *systems.*

Example 1. *Intervals are important subsets of* $\mathbb{R}$*. An open interval of* $\mathbb{R}$*, denoted by* (a, b)*, is the set* $\{x \in \mathbb{R} : a < x < b\}$*, a closed interval, denoted by* $[a, b]$ *is the set* $\{x \in \mathbb{R} : a \leq x \leq b\}$ *and a half-open interval is a set* $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ *or* $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$*. Intervals are said to be finite if* $a, b \in \mathbb{R}$ *and infinite if* $a = -\infty$ *or* $b = \infty$*. The meaning of the qualifiers 'open', 'closed' or 'half-open' as well as the meaning of the symbols* $-\infty$ *and* ∞ *will be discussed later in these notes.*

The union of sets A and B is denoted by $A \cup B$ and $A \cup B = \{x : x \in A \text{ or } x \in B\}$. The intersection of sets A and B is denoted by $A \cap B$ and $A \cap B = \{x : x \in A \text{ and } x \in B\}$. The sets A and B are disjoint if $A \cap B = \emptyset$. The difference of sets A and B is denoted by $A - B$ and $A - B = \{x : x \in A \text{ and } x \notin B\}$. The set that contains the elements that are not in A is called the complement of A and is denoted by A^c. The set $U = A \cup A^c$ is called the universal set and contains all elements that are under consideration. The symmetric difference of sets A and B is denoted by $A \Delta B$ and $A \Delta B = (A - B) \cup (B - A)$.

The next theorem shows that unions and intersections have commutative, associative and distributive properties. For a visual understanding of these properties it is useful to use of Venn diagrams.

Theorem 1. *1. (Commutative property)* $A \cap B = B \cap A$ *and* $A \cup B = B \cup A$*;*

2. *(Associative property)* $(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$ *and* $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$*;*

3. *(Distributive property)* $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ *and* $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$*.*

Proof. Left as an exercise. Note that if $A = B$ it must be that $A \subseteq B$ and $B \subseteq A$. $\square$

Remark 1. *The results in these notes are normally enunciated as 'If statement A is true then statement B is true' or $A \implies B$, that should be read as "A implies B." If it is also true that $B \implies A$, we say that A and B are equivalent, and we write $A \iff B$. The contrapositive of $A \implies B$ is 'not $B \implies$ not A' and these implications are equivalent. Hence, the veracity of a result is established either directly or by its contrapositive. In certain occasions, the veracity of a statement is established by assuming that A and 'not B' hold and then arriving at a statement known to be false or absurd, leading to the conclusion that A must imply B (see footnote 1 as an example). This method of establishing a result is known as 'proof by contradiction.'*

In some instances 'mathematical induction' can be used to establish the general veracity of propositions that we know to be true in special cases. This method of proof is justified by the following theorem.

Theorem 2. *Given a statement $P(n)$ formulated for $n \in \mathbb{N}$, suppose that:*

1. *$P(1)$ is true,*

2. *$P(k)$ true for all $k \leq n$ implies $P(n+1)$ is true.*

Then, $P(n)$ is true for all $n \in \mathbb{N}$.

Proof. Suppose $P(n)$ is not true for some n and let n_1 be the smallest natural number for which the statement P is not true. Clearly, $n_1 > 1$, since $P(1)$ is true and $n_1 - 1 \in \mathbb{N}$. $P(n)$ is valid for all $k \leq n_1 - 1$ but not n_1, contradicting 2. $\square$

It is often possible and useful to numerically index the elements of a set. For example, if a set $S = \{\text{apple}, \text{orange}, \text{banana}\}$ we can write $S = \{s_i\}_{i \in I}$ where $I = \{1, 2, 3\}$, $s_1 = \text{apple}$, $s_2 = \text{orange}$ and $s_3 = \text{banana}$. In this case, the set has a finite number of elements and the associated index set is correspondingly finite. In other cases it may be necessary to consider larger index sets such as $I = \mathbb{N}$ or $I = \mathbb{R}$.

Using an arbitrary index set I, we can represent a system $\mathcal{F}$ associated with an arbitrary set A as $\mathcal{F} = \{A_i \subset A : i \in I\}$. The next theorem is useful when manipulating sets.

Theorem 3. *(De Morgan's laws) Let I be an index set and $\mathcal{F} = \{A_i : i \in I\}$. Then,*

1. $\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c$,

2. $\left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c$.

Proof. 1. $x \in \left(\bigcup_{i \in I} A_i\right)^c \implies x \notin \bigcup_{i \in I} A_i$. But this implies that $x \notin A_i$ for all $i \in I$. Consequently, $x \in A_i^c$ for all i. That is, $x \in \bigcap_{i \in I} A_i^c$. Conversely, $x \in \bigcap_{i \in I} A_i^c$ implies $x \in A_i^c$ for all $i \in I$. Consequently, $x \notin A_i$ for all i and therefore $x \notin \bigcup_{i \in I} A_i$, which implies that $\left(\bigcup_{i \in I} A_i\right)^c$. The proof of 2. is left as an exercise. □

We now define the cartesian product of a finite collection of sets. The cartesian product will be used to define relations and functions.

Definition 1. *The cartesian product of sets A_1 and A_2, denoted by $A_1 \times A_2$, is the set of all elements given by (a_1, a_2) where $a_1 \in A_1$ and $a_2 \in A_2$. That is,*

$$A_1 \times A_2 = \{(a_1, a_2) : a_1 \in A_1 \text{ and } a_2 \in A_2\}.$$

The element (a_1, a_2) is called an ordered pair. For a finite collection of sets $\{A_1, \cdots, A_n\}$ where $n \in \mathbb{N}$, the cartesian product is given by

$$\times_{i=1}^{n} A_i = \{(a_1, \cdots, a_n) : a_i \in A_i \text{ for } i = 1, \cdots, n \text{ with } n \in \mathbb{N}\},$$

and $(a_1, \cdots, a_n)$ *is called an ordered* n*-tuple.*

Example 2. *Let* $i = 1, 2, \cdots, n$, $n \in \mathbb{N}$ *and* $A_i = \mathbb{R}$. *Then,*

$$\times_{i=1}^{n} A_i = \times_{i=1}^{n} \mathbb{R} := \mathbb{R}^n = \{(a_1, \cdots, a_n) : a_i \in \mathbb{R} \text{ for } i = 1, \cdots, n \text{ with } n \in \mathbb{N}\}.$$

What matters in identifying an n-tuple is that it is an *ordered* collection of elements, each coming from a specific set. The exact notation adopted to represent the ordered n-tuple is a matter of convenience. As such, the following representations will generally be taken as equivalent

$$(a_1, a_2, \cdots, a_n), \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}, \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}.$$

In some instances, it will be convenient or necessary to emphasize that the n-tuple is a vertical or horizontal array. In this context, if we set

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \text{ then } \mathbf{a}^T = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}.$$

$\mathbf{a}^T$ is called the transpose of $\mathbf{a}$ and we set $(\mathbf{a}^T)^T = \mathbf{a}$.

2 Relations and functions

Definition 2. *1. A relation* R *between* A *and* B *is a subset of* $A \times B$. *2. The inverse relation, denoted by* R^{-1}, *is the set with elements* $(b, a) \in B \times A$ *such that* $(a, b) \in R$.

For $a \in A$, we define the image of a under R as the set $R(a) = \{b : (a, b) \in R\}$. Note that $R(a)$ can be the empty set and that $R(a) \subseteq B$. The image of $X \subseteq A$ under R is the set $\{b : a \in X \text{ and } (a, b) \in R\} = \bigcup_{a \in X} R(a)$. For any $b \in B$, we define the inverse image of b under R as $R^{-1}(b) = \{a : (a, b) \in R\}$, which can be the empty set and $R^{-1}(b) \subseteq A$. The inverse image of $Y \subseteq B$ under R is the set $\{a : b \in Y \text{ and } (a,b) \in R\} = \bigcup_{b \in Y} R^{-1}(b)$. The image of A under R is called the range of R, and the inverse image of B under R is called the domain of R.

When R is a relation between A and B we often write $R : A \twoheadrightarrow B$, and when $(a,b) \in R$ we sometimes write aRb and say that a is related to b by R or that a and b are related by R.

Definition 3. *If $R : A \twoheadrightarrow B$ and $S : B \twoheadrightarrow C$, the composition of R and S, denoted by $R \circ S$, is the set $R \circ S = \{(a,c) : \text{there exists } b \in B \text{ with } (a,b) \in R \text{ and } (b,c) \in S\}$. In this case we write $R \circ S : A \twoheadrightarrow C$.*

Definition 4. *A function f between A and B, or from A to B, is a relation between A and B such that for every $a \in A$, $(a,b) \in f$ and $(a,c) \in f \implies b = c$.*

When f is a function from A to B we write $f : A \to B$. Usually $A = f^{-1}(B)$ and in this case A is the domain of f. When afb we write $b = f(a)$. The set B is called the co-domain of f. $G_f = \{(x,y) : \text{such that } x \in A \text{ and } y = f(x)\} \subseteq A \times B$ is usually called the graph of f.

For $X \subseteq A$, $f(X) = \{y \in B : y = f(x) \text{ for } x \in X\} \subseteq B$ is called the image of X under f and $f(A)$ is called the range of f. For $Y \subseteq B$, the inverse image of Y under f is $f^{-1}(Y) = \{a \in A : f(a) \in Y\}$.

It should be clear that f^{-1} may not be a function on B. If it is a function, i.e., if for every $y \in B$ we have that $f^{-1}(y)$ contains one, and only one element, it is called the inverse function of f. We say that $f = g$ if f and g have the same domain A and for every $a \in A$, $f(a) = g(a)$.

Example 3. *Let $A = [-1,1] \subset \mathbb{R}$ and $f = \{(x,y) : x \in A \text{ and } y = x\}$. f is clearly a function, since for every $x \in A$ there is one, and only one y satisfying the relation. The relation $R = \{(x,y) : x \in A \text{ and } y^2 = 1 - x^2\}$ is not a function, since for $x \in A$ both y and*

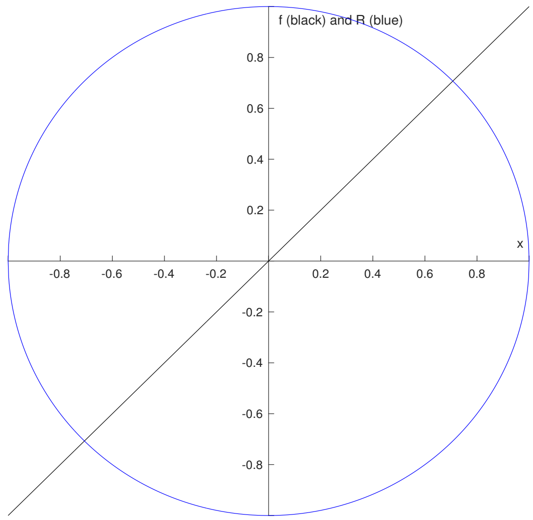

Figure 1: The function $y = x$, in black, and the relation $y^2 = 1 - x^2$, in blue

$-y$ satisfy the relation, that is $(x, y), (x, -y) \in R$. Figure 1 gives a geometric representation of these relations.

Example 4. *The indicator function of a set $A \subset \Omega$ is denoted by $I_A : \Omega \to \mathbb{R}$ and given by*

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \in A^c \end{cases},$$

where $\omega \in \Omega$. It is easily shown that $A \subseteq B \implies I_A(\omega) \leq I_B(\omega)$ for all $\omega \in \Omega$. Also, $I_A(\omega) \leq I_B(\omega)$ for all $\omega \in \Omega \implies A \subseteq B$. Hence,

$$A \subseteq B \iff I_A(\omega) \leq I_B(\omega).$$

Also, since $\omega \notin A^c \implies \omega \in A$ and $\omega \in A^c \implies \omega \notin A$, $I_{A^c}(\omega) = 1 - I_A(\omega)$ for all $\omega \in \Omega$.

Definition 5. *Let $f : A \to B$.*

1. *f is surjective or 'onto B' if $f(A) = B$,*

2. *f is injective or 'one-to-one' if for all $a, \alpha \in A$ with $a \neq \alpha$, $f(a) \neq f(\alpha)$,*

3. *f is bijective if f is onto B and one-to-one.*

The next theorem shows that inverse images, unions, intersections and differences of sets are interchangeable.

Theorem 4. *Let $f : A \to B$, I an index set and $\mathcal{F} = \{B_i : i \in I\}$ a collection of subsets of B. Then,*

1. $f^{-1}\left(\bigcup_{i \in I} B_i\right) = \bigcup_{i \in I} f^{-1}(B_i)$,

2. $f^{-1}\left(\bigcap_{i \in I} B_i\right) = \bigcap_{i \in I} f^{-1}(B_i)$,

3. $f^{-1}(B_i - B_j) = f^{-1}(B_i) - f^{-1}(B_j)$.

Proof. 1. $x \in f^{-1}\left(\bigcup_{i \in I} B_i\right) \implies f(x) \in \bigcup_{i \in I} B_i \implies f(x) \in B_i$ for some i. Hence, for some i we have $f(x) \in B_i$ and $x \in f^{-1}(B_i)$. Hence, $x \in \bigcup_{i \in I} f^{-1}(B_i)$. It is clear that all implications can be reversed, which establishes the equality of the two sets.
2. Left as an exercise, following the steps in 1.
3. $x \in f^{-1}(B_i - B_j) \implies f(x) \in B_i - B_j$. Consequently, $f(x) \in B_i$ and $f(x) \notin B_j$. Thus, $x \in f^{-1}(B_i)$ and $x \notin f^{-1}(B_j)$. Hence, $x \in f^{-1}(B_i) - f^{-1}(B_j)$. It can be easily verified that all reverse implications hold. □

The next two theorems address conditions under which images, unions, intersections and differences of sets are interchangeable.

Theorem 5. *Let $f : A \to B$, I be an index set and $\{A_i : i \in I\}$ a collection of subsets of A. Then,*

1. $f\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f(A_i)$,

2. $f\left(\bigcap_{i\in I} A_i\right) \subseteq \bigcap_{i\in I} f(A_i)$.

Proof. 1. $y \in f\left(\bigcup_{i\in I} A_i\right)$ implies that there exists $x \in \bigcup_{i\in I} A_i$ such that $(x, y) \in f$. This implies that $x \in A_i$ for some i. Hence, for some i we have $(x, y) \in f$ and $y \in f(A_i)$. Consequently, $y \in \bigcup_{i\in I} f(A_i)$. It is clear that all implications can be reversed, which establishes the equality of the two sets. 2. $y \in f\left(\bigcap_{i\in I} A_i\right)$ implies that there exists $x \in \bigcap_{i\in I} A_i$ such that $(x, y) \in f$. This implies that for all A_i, there exists an x such that $(x, y) \in f$, and consequently $y \in f(A_i)$ for all i. This implies that $y \in \bigcap_{i\in I} f(A_i)$. The reverse is not true because $y \in \bigcap_{i\in I} f(A_i)$ does not imply that there exists an $x \in \bigcap_{i\in I} A_i$ such that $(x, y) \in f$. □

Theorem 6. *Let $f : A \to B$, I be an index set and $\{A_i : i \in I\}$ a collection of subsets of A. Then,*

1. f *is one-to-one* $\iff f\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f(A_i)$,

2. f *is one-to-one* $\iff f(A - A_1) = f(A) - f(A_1)$.

Proof. 1. We start by showing that f one-to-one $\implies f\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f(A_i)$. By the definition of the image of a function, $f\left(\bigcap_{i\in I} A_i\right) \subseteq f(A_i)$ for all i. Thus, $f\left(\bigcap_{i\in I} A_i\right) \subseteq \bigcap_{i\in I} f(A_i)$. Now, if $y \in \bigcap_{i\in I} f(A_i)$ then for all i there exists $x_i \in A_i$ such that $f(x_i) = y$. Since f is one-to-one, all x_i must be the same, say x_1. Thus, $x_1 \in \bigcap_{i\in I} A_i$ and consequently $y \in f\left(\bigcap_{i\in I} A_i\right)$. Hence, $\bigcap_{i\in I} f(A_i) \subseteq f\left(\bigcap_{i\in I} A_i\right)$ and we conclude that $f\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f(A_i)$. We now show that $f\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f(A_i) \implies f$ one-to-one. Let $x_1, x_2 \in A$ such that $f(x_1) = f(x_2)$ and set $A_1 = \{x_1\}$ and $A_2 = \{x_2\}$. Then, $f(\{x_1\}) \bigcap f(\{x_2\}) \neq \emptyset$ and by assumption $f(\{x_1\}) \bigcap f(\{x_2\}) = f(\{x_1\} \bigcap \{x_2\})$ which is nonempty if and only if $x_1 = x_2$. Hence, f is one-to-one.

2. Assume that f is one-to-one. $f(x) \in f(A - A_1) \implies x \in A - A_1$. Since f is one-to-one $f(x) \neq f(x')$ for all $x' \in A_1$. Hence, $f(x) \notin f(A_1)$ and $f(x) \in f(A) - f(A_1)$ or

$f(A - A_1) \subseteq f(A) - f(A_1)$. If $f(x) \in f(A) - f(A_1)$ then $f(x) \notin f(A_1)$ which implies $x \notin A_1$ and $x \in A - A_1$. Consequently $f(x) \in f(A - A_1)$ and $f(A) - f(A_1) \subseteq f(A - A_1)$ (one-to-one property of f not used). Thus, $f(A) - f(A_1) = f(A - A_1)$.

Now, assume that $f(A - A_1) = f(A) - f(A_1)$. Let $A_1 = \{x\}$ and consider $x' \neq x$. Clearly, $x' \in A - A_1$ and $f(x') \in f(A - A_1)$. But by assumption $f(x') \in f(A) - f(\{x\})$. Hence, $f(x') \notin f(\{x\})$ and $f(x') \neq f(x)$, establishing that f is one-to-one. □

3 The limit of a sequence of sets

Often, it is necessary to use the infinity symbols $-\infty$ or ∞ in calculations. In these cases we work with the extended real line, i.e., $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\} := [-\infty, \infty]$. When we write $x \in \mathbb{R}$ we mean $-\infty < x < \infty$. The extended real line inherits the ordering as well as the rules for addition and multiplication we associate with $\mathbb{R}$. These rules are augmented as follows in $\bar{\mathbb{R}}$:

1. $x + \infty = \infty$, $x + (-\infty) = -\infty$, for $x \in \mathbb{R}$
2. $\infty + \infty = \infty$, $-\infty + (-\infty) = -\infty$
3. $-\infty + \infty$ and $\frac{\pm\infty}{\pm\infty}$ are not defined
4. $0 \times \infty = 0$, $0 \times -\infty = 0$
5. $\pm x \times \infty = \pm\infty$, $\pm x \times (-\infty) = \mp\infty$, for $x \in \mathbb{R}$
6. $\infty \times \infty = \infty$, $\infty \times -\infty = -\infty$, $-\infty \times -\infty = \infty$

Functions that take values in $\bar{\mathbb{R}}$ are called *numerical* functions.

Definition 6. *Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of arbitrary sets and define the following: $I_1 = \bigcap_{n\geq 1} A_n$, $I_2 = \bigcap_{n\geq 2} A_n$, $\cdots$ and $U_1 = \bigcup_{n\geq 1} A_n$, $U_2 = \bigcup_{n\geq 2} A_n$, $\cdots$. Note that $I_1 \subseteq I_2 \subseteq \cdots$ and $U_1 \supseteq$*

$U_2 \supseteq \cdots$. *We write* $\liminf_{n\to\infty} A_n = \bigcup_{k\in\mathbb{N}} I_k$ *and* $\limsup_{n\to\infty} A_n = \bigcap_{k\in\mathbb{N}} U_k$. *If* $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$ *we say that the collection* $\{A_n\}_{n\in\mathbb{N}}$ *has a limit and we write*

$$A := \lim_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n.$$

Example 5. *a) Let* $A_n = [0, n/(n+1))$ *for* $n \in \mathbb{N}$. *Then,* $U_1 = \bigcup_{n=1}^{\infty} A_n = [0,1)$, $U_2 = [0,1)$, $\cdots$. $I_1 = \bigcap_{n=1}^{\infty} A_n = [0,1/2)$, $I_2 = \bigcap_{n=2}^{\infty} A_n = [0,2/3)$, $\cdots$. *Hence,* $\liminf_{n\to\infty} A_n = \bigcup_{k\in\mathbb{N}} I_k = [0,1)$ *and* $\limsup_{n\to\infty} A_n = \bigcap_{k\in\mathbb{N}} U_k = [0,1)$.
b) Let $A_n = (0,1/n)$ *for* $n \in \mathbb{N}$. *Then,* $U_1 = \bigcup_{n=1}^{\infty} A_n = (0,1)$, $U_2 = \bigcup_{n=2}^{\infty} A_n = (0,1/2)$, $\cdots$ *and* $I_1 = \bigcap_{n=1}^{\infty} A_n = \emptyset$, $I_2 = \bigcap_{n=2}^{\infty} A_n = \emptyset$, $\cdots$. *Hence,* $\liminf_{n\to\infty} A_n = \bigcup_{k\in\mathbb{N}} I_k = \emptyset$ *and* $\limsup_{n\to\infty} A_n = \bigcap_{k\in\mathbb{N}} U_k = \emptyset$.

The next theorem provides a characterization for $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$ using the indicator function. Here, $x \notin \mathbb{R}$ means that $x \in \overline{\mathbb{R}} - \mathbb{R}$.

Theorem 7. *Let* $\{A_n\}_{n\in\mathbb{N}}$ *be a collection of subsets of* A. *Then,*

$$1. \liminf_{n\to\infty} A_n = \left\{ x : \sum_{n\in\mathbb{N}} I_{A_n^c}(x) < \infty \right\},$$

$$2. \limsup_{n\to\infty} A_n = \left\{ x : \sum_{n\in\mathbb{N}} I_{A_n}(x) = \infty \right\}.$$

Proof. 1. $x \in \liminf_{n\to\infty} A_n \implies x \in \bigcap_{n\geq k} A_n$ for some $k \in \mathbb{N}$. But by DeMorgan's Laws $\bigcap_{n\geq k} A_n = \left(\bigcup_{n\geq k} A_n^c \right)^c$. Hence, $x \notin \bigcup_{n\geq k} A_n^c$ and consequently $x \notin A_n^c$ for all $n \geq k$. Thus, $I_{A_n^c}(x) = 0$ for all $n \geq k$. Hence,

$$\sum_{n\in\mathbb{N}} I_{A_n^c}(x) = \sum_{n<k} I_{A_n^c}(x) + \sum_{n\geq k} I_{A_n^c}(x) = \sum_{n<k} I_{A_n^c}(x) < \infty.$$

Thus, $\liminf_{n\to\infty} A_n \subseteq \{x : \sum_{n\in\mathbb{N}} I_{A_n^c}(x) < \infty\}$.

Now, $x \in \{x : \sum_{n\in\mathbb{N}} I_{A_n^c}(x) < \infty\}$ implies that x belongs to a finite number of A_n^c. That is, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $I_{A_n^c}(x) = 0$. That is, $x \notin A_n^c$ for all

$n \geq n_0$. But this implies that $x \notin \bigcup_{n \geq n_0} A_n^c$ which implies $x \in \left(\bigcup_{n \geq n_0} A_n^c\right)^c = \bigcap_{n \geq n_0} A_n$. By definition, this means that $x \in \liminf\limits_{n \to \infty} A_n$. Thus, $\liminf\limits_{n \to \infty} A_n = \left\{x : \sum\limits_{n \in \mathbb{N}} I_{A_n^c}(x) < \infty\right\}$.

2. $x \in \limsup\limits_{n \to \infty} A_n \implies x \in \bigcup_{n \geq k} A_n$ for all $k \in \mathbb{N}$. But, by De Morgan's Laws $\bigcup_{n \geq k} A_n = \left(\bigcap_{n \geq k} A_n^c\right)^c$. Hence, $x \notin \bigcap_{n \geq k}^{\infty} A_n^c$ and it must be that $x \in A_n$ for some $n \geq k$ and all $k \in \mathbb{N}$. Consequently, $I_{A_n}(x) = 1$ for some $n \geq k$ and all $k \in \mathbb{N}$. Hence, $x \in \left\{x : \sum\limits_{n \in \mathbb{N}} I_{A_n}(x) = \infty\right\}$. Conversely, if $x \in \left\{x : \sum\limits_{n \in \mathbb{N}} I_{A_n}(x) = \infty\right\}$, then there are only finitely many A_n that do not contain x. Hence, for all $k \in \mathbb{N}$ we have $x \in \bigcup_{n \geq k} A_n$, which implies that $x \in \bigcap\limits_{k \in \mathbb{N}} \bigcup\limits_{n \geq k} A_n = \limsup\limits_{n \to \infty} A_n$. □

Remark 2. *1. Because of the characterization given in Theorem 7, it is common to refer to* $\liminf\limits_{n \to \infty} A_n$ *as the set of* x*'s for which* $x \in A_n$ *for all* n *except a finite number* ($\sum_{n \in \mathbb{N}} I_{A_n^c}(x) < \infty$)*. And to refer to* $\limsup\limits_{n \to \infty} A_n$ *is the set of* x*'s for which* $x \in A_n$ *infinitely often (i.o).*

2. Note that $x \in \liminf\limits_{n \to \infty} A_n \implies x \in \bigcup\limits_{k \in \mathbb{N}} I_k \implies x \in \bigcap\limits_{n \geq k} A_n$ *for some* k*. Hence,* $x \in A_n$ *for all* $n \geq$ *some* k*. Hence,* $x \in \bigcup_{n \geq k} A_n$ *for all* k*, and consequently, by definition* $x \in \limsup\limits_{n \to \infty} A_n$*. Thus,*

$$\liminf_{n \to \infty} A_n \subseteq \limsup_{n \to \infty} A_n.$$

3. By De Morgan's Laws $\left(\liminf\limits_{n \to \infty} A_n\right)^c = \limsup\limits_{n \to \infty} A_n^c$.

4 Cardinality of sets

The number of elements in a set, or its cardinality, is an important concept. When sets have finitely many elements, we can obtain a count of these elements and, at least conceptually, establish its cardinality. The cardinality of sets that have infinitely many elements cannot be established in this simple manner. We start with a definition of equality of the cardinality of two sets.

Definition 7. *Sets* A *and* B *have the same cardinality, and we write* $\#A = \#B$*, if there* *is a function $f : A \to B$ that is one-to-one and onto. We write $\#A \leq \#B$ if there is a one-to-one function $g : A \to B$ (not necessarily onto).*

If $B = \mathbb{N}$ and $\#A \leq \#\mathbb{N}$ then every element of A can be associated with a unique element from $\mathbb{N}$. If $\#A < \#B$, then there exists no $g : A \to B$ that is one-to-one and onto.

Definition 8. *If $\#A \leq \#\mathbb{N}$ we say that A is countable. The cardinality of $\mathbb{N}$ is denoted by $\aleph_0$ (aleph naught). If $\#A > \aleph_0$, we say that A is uncountable.*

We can establish the cardinality of some important sets by defining suitable bijective functions.

Example 6. *1. Consider $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ and the function $f : \mathbb{Z} \to \mathbb{N}$ with*

$$f(z) = \begin{cases} 2z, & \text{if } z > 0 \\ 2|z| + 1, & \text{if } z \leq 0 \end{cases} .$$

f is clearly bijective. Hence, $\#\mathbb{Z} = \#\mathbb{N}$ and we conclude that there are as many integers as natural numbers.

2. Let $A = \mathbb{N} \times \mathbb{N}$ and denote an arbitrary element of A by $a = (a_1, a_2)$. Consider the function $f : A \to \mathbb{N}$ given by

$$f(a) = \frac{(a_1 + a_2)(a_1 + a_2 - 1)}{2} - a_2 + 1.$$

f is clearly a bijective. Hence, $\#\mathbb{N} \times \mathbb{N} = \#\mathbb{N}$ and we conclude that the cartesian product of two sets of natural numbers has the same cardinality of the natural numbers. See Figure 2 for a diagrammatic representation of how f counts the elements of $\mathbb{N} \times \mathbb{N}$.

The next theorem shows that countable unions of countable sets are countable sets.

Theorem 8. *Let $\{A_n\}_{n \in \mathbb{N}}$ be a collection of sets (as written, it is countable) with each A_n countable. Then, $A = \bigcup_{n \in \mathbb{N}} A_n$ is countable.*

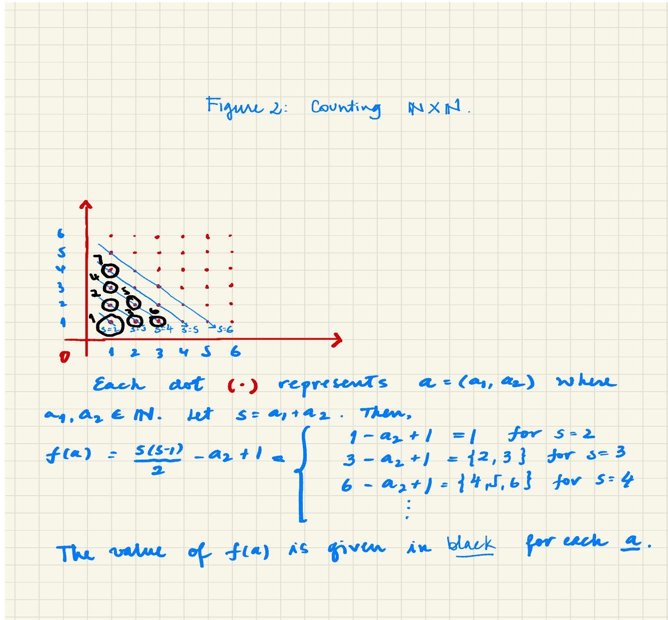

Figure 2: Counting $\mathbb{N} \times \mathbb{N}$

Proof. Since each A_n is countable, we can enumerate its elements as $A_n = \{A_{n,1}, A_{n,2}, \cdots\}$. Hence, $A = \{A_{1,1}, A_{1,2}, A_{1,3}, \cdots, A_{2,1}, A_{2,2}, A_{2,3}, \cdots\} = \{A_{i,j} : (i, j) \in \mathbb{N} \times \mathbb{N}\}$, which is countable by item 2 in the previous Example. □

This theorem can be used to establish that $\mathbb{Q}$, the set of rational numbers, is countable. To verify this, note that if $q \in \mathbb{Q}$ then $q = a_1/a_2$ where $a_1, a_2 \in \mathbb{Z}$ with $a_2 \neq 0$. Now, consider $q > 0$ and note that in this case $q = a_1/a_2$ can be written using the pairs (a_1, a_2) in item 2. Similarly, $q < 0$ is countable. Since Q is the union of positive and negative rationals with 0, we conclude that Q is countable. A formal proof is left as an exercise.

Theorem 9. *For any set A, $\#A < \#2^A$.*

Proof. We need to show that any $g : A \to 2^A$ that is one-to-one is not onto. Let $B = \{a \in A : a \notin g(a)\}$ be the elements of A whose images do not contain a. Note that in this case the images are subsets of A. Since $B \in 2^A$, if g is onto, there exists $a' \in A$ such that $g(a') = B$. Hence, $a' \in B \iff a' \notin g(a') \iff a' \notin B$ which is impossible. □

Theorem 10. *(Cantor-Bernstein) If $\#A \leq \#B$ and $\#B \leq \#A$, then $\#A = \#B$.*

Proof. By assumption, there exists $f : A \to B$ and $g : B \to A$ that are both one-to-one. Since $g : B \to g(B)$ is bijective, $\#B = \#g(B)$ and the theorem will be proved if we show that $\#A = \#g(B)$. As $g(B) \subseteq A$, we can take g to be the identity function, i.e., $g(b) = b$ for all $b \in B$, and proceed with $B \subseteq A$. We must obtain a bijection $h : A \to B$.

Put $A_0 = A$ and define $A_1 := f(A) \subseteq B$, $A_2 := f(f(A)) = f^2(A) = f(A_1)$, $A_3 = f^3(A) = f(A_2), \cdots$ with f^0 being the identity function. Then, we write $A_i = f^i(A)$ for $i = 0, 1, \cdots$. Since $B \subseteq A$ we put $B_0 = B$ and similarly define $B_1 := f(B) \subseteq B$, $B_2 := f(f(B)) = f^2(B) = f(B_1)$, $B_3 = f^3(B) = f(B_2), \cdots$ and write $B_i = f^i(B)$ for $i = 0, 1, \cdots$. Now, since $f(A) \subseteq B$ and $B \subseteq A$,

$$f^{i+1}(A) = f^i(f(A)) \subseteq f^i(B) \subset f^i(A) \text{ or } A_{i+1} \subseteq B_i \subseteq A_i \text{ for } i = 0, 1, \cdots$$

Hence, we have $A \supseteq B \supseteq A_1 \supseteq B_1 \supseteq A_2 \supseteq \cdots$ Now, define a function $h : A \to B$ by

$$h(x) := \begin{cases} f(x) & \text{if } x \in \bigcup_{i \in \mathbb{N}_0} (A_i - B_i), \\ x & \text{if } x \notin \bigcup_{i \in \mathbb{N}_0} (A_i - B_i). \end{cases}$$

Now,

$$
\begin{aligned}
h(A) &= h\left(\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)\bigcup\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)^c\right)\\
&= h\left(\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)\right)\bigcup h\left(\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)^c\right)\\
&= f\left(\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)\right)\bigcup\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)^c\\
&= \left(\bigcup_{i\in\mathbb{N}_0}(f(A_i) - f(B_i))\right)\bigcup\left(\bigcup_{i\in\mathbb{N}_0}(A_i - B_i)\right)^c \text{ since } f \text{ is injective.}\\
&= \left(\bigcup_{i\in\mathbb{N}_0}(A_{i+1} - B_{i+1})\right)\bigcup\left((A-B)\bigcup\bigcup_{i\in\mathbb{N}}(A_i - B_i)\right)^c\\
&= \left(\bigcup_{i\in\mathbb{N}_0}(A_{i+1} - B_{i+1})\right)\bigcup\left((A-B)^c\bigcap\left(\bigcup_{i\in\mathbb{N}_0}(A_{i+1} - B_{i+1})\right)^c\right) = C\bigcup((A-B)^c\bigcap C^c)
\end{aligned}
$$

by letting $C := \bigcup_{i\in\mathbb{N}_0}(A_{i+1} - B_{i+1})$.

But $C\cup((A-B)^c\cap C^c) = C\cup((A^c\cup B)\cap C^c) = (C\cup A^c\cup B)\cap(C\cup C^c) = (C\cup A^c\cup B)\cap A$. But $C\subseteq B$, hence $C\cup B = B$ and $C\cup((A-B)^c\cap C^c) = (B\cup A^c)\cap A = A\cap B = B$ since $B\subseteq A$. Hence, $h(A) = B$ and h is onto.

We now show that h is one-to-one. To this end, let a, $a'\in A$ and let $h(a) = h(a')$. We must show that $a = a'$. There are the following cases:

1. $a, a' \in \bigcup_{i\in\mathbb{N}_0}(A_i - B_i)$. Then, $h(a) = f(a)$ and $h(a') = f(a')$ and consequently $f(a) = f(a')$. But since f is one-to-one, $a = a'$;

2. $a, a' \notin \bigcup_{i\in\mathbb{N}_0}(A_i - B_i)$.Then, $a = h(a) = h(a') = a'$ since f;

3. $a \in \bigcup_{i\in\mathbb{N}_0}(A_i - B_i)$ and $a' \notin \bigcup_{i\in\mathbb{N}_0}(A_i - B_i)$ which implies $a' \notin (A_i - B_i)$ for all $i\in\mathbb{N}_0$. Then, $f(a) = h(a) = h(a') = a'$. Now, $a' = f(a)$ implies $a' \in f(A_i - B_i)$ for some i, but since f is one-to-one $f(A_i - B_i) = f(A_i) - f(B_i) = A_{i+1} - B_{i+1}$. Hence, $a' \in A_{i+1} - B_{i+1}$ which is ruled out by assumption. Hence, this case cannot occur.

4. $a' \in \bigcup_{i \in \mathbb{N}_0} (A_i - B_i)$ and $a \notin \bigcup_{i \in \mathbb{N}_0} (A_i - B_i)$ also cannot occur given the arguments we made for case 3.

Hence, h is one-to-one, and since we have already established that it is onto, it is bijective, completing the proof. $\square$

Theorem 11. *If $A \subseteq B$ then $\#A \leq \#B$.*

Proof. Since $A \subseteq B$, $f : A \to B$ such that $f(a) = a$ is one-to-one. Thus, $\#A \leq \#B$. $\square$

A direct consequence of the last theorem is that subsets of countable sets are countable. Not all sets are countable. The next theorem shows that the interval $(0, 1)$ is uncountable. We use this result to say that the set of real number $\mathbb{R}$ is also uncountable.

Theorem 12. *The interval $(0, 1)$ is uncountable and its cardinality $\#(0, 1) := \mathcal{C}$ is called the continuum.*

Proof. Every $x \in (0, 1)$ can be written as $0.x_1x_2, \cdots$ where $x_i \in \{0, 1, \cdots, 9\}$. However, this representation is not unique since, e.g., $0.500\cdots = 0.499\cdots$. To avoid non-uniqueness, whenever $x = 0.x_1x_2, \cdots, x_n$ where n is a natural number (finite number of decimal points), we substitute x_n with $x_n - 1$ and set all subsequent digits to 9. As a result, every $x \in (0, 1)$ is now represented uniquely by an infinite decimal expansion. The proof, follows by contradiction. Assume, $(0, 1)$ is countable and let all its elements be represented by the enumeration $\{y_1, y_2, \cdots\}$. Then,

$$y_1 = 0.y_{11}y_{12}y_{13}\cdots$$
$$y_2 = 0.y_{21}y_{22}y_{23}\cdots$$
$$\vdots$$

Now, consider $x = 0.x_1x_2, \cdots$ where x_1 is any number in $\{0, 1, \cdots, 9\}$ different from y_{11}, x_2 is any number in $\{0, 1, \cdots, 9\}$ different from y_{22}, and so on. Hence, $x \neq y_i$ for all $i = 1, 2, \cdots$. Consequently, $\{y_1, y_2, \cdots\}$ cannot include all $x \in (0, 1)$. $\square$

Theorem 13. $\#\mathbb{R} = \#(0,1) = \mathcal{C}$.

Proof. Let $f : (0,1) \to \mathbb{R}$ with $f(x) = (1-x)^{-1} - x^{-1}$. We can promptly verify that $f((0,1)) = \mathbb{R}$, i.e., onto and f is strictly increasing on $(0,1)$, hence one-to-one. □

In the next section we discuss some of the properties of $\mathbb{R}$ and its structure.

5 The real numbers

We assume the existence of two operations that can be performed on the elements of $\mathbb{R}$. The first operation is called the sum of any two $x, y \in \mathbb{R}$ and is denoted by $x + y$. The second operation is called the (scalar) multiplication of any two $x, y \in \mathbb{R}$ and is denoted by xy. It is assumed that xy, $x + y \in \mathbb{R}$. For any $x, y, z \in \mathbb{R}$, these operations are assumed to satisfy the following axioms (field):[3]

1. $x + y = y + x\,; xy = yx$ (commutative property),

2. $(x + y) + z = x + (y + z)\,; (xy)z = x(yz)$ (associative property),

3. $(x + y)z = xz + yz$ (distributive property),

4. For any x, y there exists z such that $x + z = y$. This z is denoted by $y - x$. If $y = 0$, then $z = -x$, which is called the negative of x. If $y = x$, $z = x - x = 0$.

5. There exists at least one $x \neq 0$. In this case, for any x, y there exists z such that $xz = y$. This z is denoted by y/x. If $y = x$, $z = 1$. If $y = 1$, then $z = 1/x$ which is called the reciprocal of x.

In addition, we assume that there is a relation, denoted by $<$, which establishes an ordering for the elements of $\mathbb{R}$. The elements of $\mathbb{R}$ satisfy the following axioms (order):

[3]Recall that in mathematics an axiom is a statement that is taken to be true, requiring no exercise to establish its validity.

1. For any x, y either $x < y$, $y < x$ or $y = x$,

2. $x < y \implies x + z < y + z$ for all z,

3. $x, y > 0 \implies xy > 0$.

4. $x > y$ and $y > z \implies x > z$.

Writing $x \leq y$ means that $x = y$ or $x < y$.

Lastly, we assume that $\mathbb{R}$ satisfies the 'completeness axiom.' Before we state this axiom we need additional definitions.

Definition 9. *$u \in \mathbb{R}$ is an upper bound for $A \subset \mathbb{R}$ if every $a \in A$ satisfies $a \leq u$. $l \in \mathbb{R}$ is a lower bound for $A \subset \mathbb{R}$ if every $a \in A$ satisfies $a \geq l$.*

If A has an upper bound, it is said to be bounded above, and if A has a lower bound it is said to be bounded below. It is evident that if u is an upper bound for A, every real number $u_1 > u$ is also an upper bound for A. If $u \in A$, u is called the maximum element of A. Similarly, if l is a lower bound for A, every real number $l_1 < l$ is also a lower bound for A. If $l \in A$, l is called the minimum element of A.

Definition 10. *1. If $A \subset \mathbb{R}$ is bounded above, $s \in \mathbb{R}$ is the least upper bound for A, or the supremum of A, if s is an upper bound for A and there is no real number $u < s$ that is an upper bound for A. We write $s = \sup A$.*
2. If $A \subset \mathbb{R}$ is bounded below, $\iota \in \mathbb{R}$ is the greatest lower bound for A, or the infimum of A, if ι is a lower bound for A and there is no real number $l > \iota$ that is a lower bound for A. We write $\iota = \inf A$.

As defined, it is easy to verify (you should try) that whenever they exist, the supremum and the infimum are unique. The *completeness axiom* states that every nonempty $A \subset \mathbb{R}$ that is bounded above has a supremum. The next theorem says that subsets of $\mathbb{R}$ that have a supremum contain elements that are arbitrarily close to the supremum.

Theorem 14. *Let $A \subset \mathbb{R}$ be nonempty and bounded above with $s = \sup A$. Then, for every $x < s$ there exists $a \in A$ such that $x < a \leq s$.*

Proof. By definition of $\sup A$, $a \leq s$ for all $a \in A$. If all $a \leq x$, then x is an upper bound for A and cannot be less than s. Hence, there must be $a > x$. □

Theorem 15. $\mathbb{N}$ *has no upper bound.*

Proof. Suppose $\mathbb{N}$ is bounded above. Then, it has a supremum $s \in \mathbb{R}$. By the previous theorem, there exists $n \in \mathbb{N}$ such that $s - 1 < n \leq s$. Then, $s < n + 1$. But $n + 1 \in \mathbb{N}$, contradicting that $s = \sup \mathbb{N}$. □

Corollary 1. *(Archimedean property) Given any $x \in \mathbb{R}$, there exists an integer $n > x$.*

Proof. If not, that is, every $n \leq x$, then x would be an upper bound for $\mathbb{N}$, contradicting the previous theorem. □

Corollary 2. *For any $x, y \in \mathbb{R}$ such that $x < y$, there exists $q \in \mathbb{Q}$ such that $x < q < y$.*

Proof. Suppose $x \geq 0$. Then $0 < y - x$ and $0 < \frac{1}{y-x}$. By the Archimedean property, there exist $n \in \mathbb{N}$ such that $\frac{1}{y-x} < n$ or, equivalently,

$$\frac{1}{n} < y - x. \tag{1}$$

Now, for this n and y, let $S_{y,n} = \{i \in \mathbb{N} : y \leq i/n\}$. Again, by Archimedean property, $S_{y,n} \neq \emptyset$. Hence, there exists a smallest element p in the set such that

$$\frac{p-1}{n} < y \leq \frac{p}{n}. \tag{2}$$

Using equations (1) and (2) we have $x = y - (y - x) \leq \frac{p}{n} - (y - x) < \frac{p-1}{n}$ and $x < \frac{p-1}{n} < y$. Since $\frac{p-1}{n} \in \mathbb{Q}$, we have the desired result.

If $x < 0$, then by the Archimedean property there exists $n \in \mathbb{N}$ such that $-x < n$ or, equivalently, $n + x > 0$. By the first part of the proof, there exists $q \in \mathbb{Q}$ such that $n + x < q < n + y$ or $x < q - n < y$. Since, $q - n \in \mathbb{Q}$ we have the desired result. □

Theorem 16. *Let $x, y \in \mathbb{R}$. If for all $\varepsilon > 0$ we have $x \leq y + \varepsilon$, then $x \leq y$.*

Proof. We will prove the contrapositive, i.e., if $x > y$ then for some $\varepsilon > 0$, $y + \varepsilon < x$. Let $\varepsilon = \frac{1}{2}(x - y) > 0$. Then, $y + \varepsilon = y + \frac{1}{2}(x - y) = \frac{1}{2}(x + y) < \frac{1}{2}(x + x) = x$. $\square$

An important function on the $\mathbb{R}$ is the absolute value function $|\cdot| : \mathbb{R} \to [0, \infty)$ given by

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

We now establish the triangle inequality.

Theorem 17. *For any $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$.*

Proof. By definition, $-|x| \leq x \leq |x|$ for all $x \in \mathbb{R}$. Hence, $-(|x| + |y|) \leq x + y \leq |x| + |y|$ and $|x + y| \leq |x| + |y|$. $\square$

6 Sequences

Sequences play important roles in mathematics and many important concepts can be characterized using their behavior or properties.

Definition 11. *1. By a finite sequence of $n \in \mathbb{N}$ terms we mean the image of a function f defined on $\{1, \cdots, n\}$. The image of f is written as $\{f(1), \cdots, f(n)\}$ or $\{f_1, \cdots, f_n\}$ or $\{f_j\}_{j=1}^n$.*
2. By an infinite sequence we mean the image of a function f defined on $\mathbb{N}$. The image of f is written as $\{f(1), f(2), \cdots\}$ or $\{f_1, f_2, \cdots\}$ or $\{f_j\}_{j=1}^\infty$ or $\{f_j\}_{j\in\mathbb{N}}$.
3. If $\{f_1, f_2, \cdots\}$ is an infinite sequence and $k : \mathbb{N} \to \mathbb{N}$ is such that

$$k(m) < k(n) \text{ whenever } m < n,$$

the composition $(f \circ k)(n) = f(k(n)) = f(k_n)$ with image $\{f_{k_1}, f_{k_2}, \cdots\} = \{f_{k_n}\}_{n\in\mathbb{N}}$ is called a subsequence of $\{f_1, f_2, \cdots\}$.

Below are examples of sequences and subsequences.

Example 7. *Let*

$$\{f(n)\}_{n\in\mathbb{N}} = \{log(n)\}_{n\in\mathbb{N}} = \{f(1), f(2), f(3), f(4), \cdots\} = \{0, 0.6931, 1.0986, 1.3863, \cdots\},$$

$$\{k(n)\}_{n\in\mathbb{N}} = \{2n\}_{n\in\mathbb{N}} = \{2, 4, \cdots\}.$$

Then, $\{(f\circ k)(n)\}_{n\in\mathbb{N}} = \{log(2n)\}_{n\in\mathbb{N}} = \{f(2), f(4), \cdots\} = \{0.6931, 1.3863, \cdots\}.$

Sequences in $\mathbb{R}$ are of particular interest. According to Definition 10, a sequence $\{x_n\}_{n\in\mathbb{N}}$ must be bounded above (below) for us to define its supremum (infimum). However, it is convenient to think of ∞ $(-\infty)$ as the supremum (infimum) of a sequence that is not bounded above (below). What follows are the definitions of the limit inferior and limit superior of a sequence of real numbers.

Definition 12. *For a sequence of real numbers* $\{x_n\}_{n\in\mathbb{N}}$*, the lower limit* $\liminf_{n\to\infty} x_n := \sup_{k\in\mathbb{N}} \inf_{n\geq k} x_n$*. The upper limit is* $\limsup_{n\to\infty} x_n := \inf_{k\in\mathbb{N}} \sup_{n\geq k} x_n$*.*

Note that $i_k = \inf_{n\geq k} x_n$ and $s_k = \sup_{n\geq k} x_n$ for $k\in\mathbb{N}$ are increasing and decreasing sequences in $\mathbb{R}$, with $i_k \in [-\infty, \infty)$ and $s_k \in (-\infty, \infty]$. Hence, $\liminf_{n\to\infty} x_n$, $\limsup_{n\to\infty} x_n \in \bar{\mathbb{R}}$.

Since $\inf_{n\geq k} x_n = -\sup_{n\geq k}(-x_n)$, we have

$$\begin{aligned}\liminf_{n\to\infty} x_n = \sup_{k\in\mathbb{N}} \inf_{n\geq k} x_n &= \sup_{k\in\mathbb{N}} \left(-\sup_{n\geq k}(-x_n)\right)\\ &= -\inf_{k\in\mathbb{N}} \sup_{n\geq k}(-x_n)\\ &= -\limsup_{n\to\infty}(-x_n).\end{aligned}$$

Also, since by definition $s_k \geq i_k$ for all $k\in\mathbb{N}$, it must be that $\limsup_{n\to\infty} x_n := \inf_{k\in\mathbb{N}} s_k \geq \sup_{k\in\mathbb{N}} i_k := \liminf_{n\to\infty} x_n$.

A sequence is said to be increasing (decreasing) if for all $n\in\mathbb{N}$, $x_n \leq x_{n+1}$ $(x_n \geq x_{n+1})$. A sequence that is increasing or decreasing is called monotonic. When the inequality on the image is strict, i.e., $x_n < x_{n+1}$ ($x_n > x_{n+1}$), we say that the sequence is strictly increasing (decreasing) or strictly monotonic.

Theorem 18. *Let* $\{x_n\}_{n\in\mathbb{N}}$ *be a sequence in* $\mathbb{R}$. *It contains a decreasing or an increasing subsequence, or both.*

Proof. Let $S = \{s \in \mathbb{N} : x_s > x_n \text{ for all } n > s\}$. This is the set of indices that are associated with members of the sequence that are larger than all subsequent members of the sequence. Note that S is either finite or infinite set. It is always the case that the set $\{x_j : j \in S\}$ is such that $x_j > x_{j'}$ if $j < j'$. In the first case, when S is finite, S is bounded and has a supremum, say $\sup S$. Now, $\sup S + 1$ is the index of a member of the sequence which is not in S. Therefore, there exists another member of the sequence (with index not in S) that is larger than $x_{\sup S+1}$, which is denoted by $x_{\sup S+2}$. Similarly, there exist $x_{\sup S+3}$ such that $x_{\sup S+2} \leq x_{\sup S+3}$. The argument can be repeated *ad infinitum* to construct an increasing subsequence of x_n. In the second case, when S is infinite, there always exists a sequence member $x_j > x_{j'}$ if $j < j'$ and $j, j' \in S$. □

The next theorem relies on the notion of the limit of a sequence. From elementary Calculus, we will say that a sequence of real numbers $\{x_n\}_{n\in\mathbb{N}}$ converges to $x \in \mathbb{R}$ if for all $\epsilon > 0$ there exists $N(\epsilon) \in \mathbb{N}$ such that for all $n > N(\epsilon)$ we have $|x_n - x| < \epsilon$. In this case we write, $\lim_{n\to\infty} x_n = x$. The notation $N(\epsilon)$ means 'N depending on ϵ' which is often also represented by N_ϵ.

It is easy to show that in the case where a limit exists, we have

$$\lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n.$$

Theorem 19. *Let* $\{x_n\}_{n\in\mathbb{N}}$ *be a monotonic sequence in* $\mathbb{R}$.

$$\{x_n\}_{n\in\mathbb{N}} \text{ converges } \iff \{x_n\}_{n\in\mathbb{N}} \text{ is bounded.}$$

Proof. (⇐) Let $\{x_n\}_{n\in\mathbb{N}}$ be bounded and increasing. Then, there exist $\bar{x} = \sup_{n\in\mathbb{N}}\{x_n\}$ such that $x_n \leq \bar{x}$ for all n. By definition of supremum, for all $\epsilon > 0$, $\bar{x} - \epsilon$ is not an upper bound for $\{x_n\}_{n\in\mathbb{N}}$. Therefore, there exists $N(\epsilon) \in \mathbb{N}$ such that for $n > N(\epsilon)$

$$\bar{x} - \epsilon < x_n \leq \bar{x} < \bar{x} + \epsilon \text{ or } |x_n - \bar{x}| < \epsilon.$$

Hence, $\{x_n\}_{n\in\mathbb{N}}$ converges to $\bar{x} = \sup_{n\in\mathbb{N}}\{x_n\}$.

Let $\{x_n\}_{n\in\mathbb{N}}$ be bounded and decreasing. Then, there exist $\underline{x} = \inf_{n\in\mathbb{N}}\{x_n\}$ such that $x_n \geq \underline{x}$ for all n. By definition of infimum, for all $\epsilon > 0$, $\underline{x} + \epsilon$ is not a lower bound for $\{x_n\}_{n\in\mathbb{N}}$. Therefore, there exists $N(\epsilon) \in \mathbb{N}$ such that for $n > N(\epsilon)$

$$\underline{x} - \epsilon < \underline{x} \leq x_n < \underline{x} + \epsilon \text{ or } |x_n - \underline{x}| < \epsilon.$$

Hence, $\{x_n\}_{n\in\mathbb{N}}$ converges to $\underline{x} = \inf_{n\in\mathbb{N}}\{x_n\}$.

(⇒) If $\{x_n\}_{n\in\mathbb{N}}$ converges, let x denote its limit. Then, for all $\epsilon > 0$ there exists $N(\epsilon) \in \mathbb{N}$ such that for $n > N(\epsilon)$

$$x - \epsilon < x_n < x + \epsilon.$$

Let $r_j = |x_j - x|$ for $j = 1, \ldots, N(\epsilon)$ and $r := \max_{1\leq j\leq N(\epsilon)} r_j$. Then for all $n \in \mathbb{N}$ we have

$$|x_n - x| \leq r + \epsilon.$$

Thus, $\{x_n\}_{n\in\mathbb{N}}$ is bounded. □

It is useful to establish the relationship between limit inferior and limit superior as defined for sets and as defined for sequence of real numbers.

Theorem 20. *Let $x \in \mathbb{X}$ the domain of an indicator function. Then,*

$$\liminf_{j\to\infty} I_{A_j}(x) = I_{\liminf_{j\to\infty} A_j}(x) \text{ and } \limsup_{j\to\infty} I_{A_j}(x) = I_{\limsup_{j\to\infty} A_j}(x).$$

Proof. Note that for any collection of sets $S_1, S_2, \dots$

$$\begin{aligned} I_{\bigcap_{k\in\mathbb{N}} S_k}(x) = 1 \quad &\iff \quad x \in \bigcap_{k\in\mathbb{N}} S_k \\ &\iff \quad x \in S_k, \text{ for all } k \iff I_{S_k} = 1 \text{ for all } k \iff \inf_{k\in\mathbb{N}} I_{S_k} = 1. \end{aligned}$$

Also,

$$\begin{aligned} I_{\bigcup_{k\in\mathbb{N}} S_k}(x) = 1 \quad &\iff \quad x \in \bigcup_{k\in\mathbb{N}} S_k \\ &\iff \quad x \in S_k, \text{ for some } k \iff I_{S_k} = 1 \text{ for some } k \iff \sup_{k\in\mathbb{N}} I_{S_k} = 1. \end{aligned}$$

Now, $I_{\liminf_{j\to\infty} A_j}(x) = I_{\bigcup_{k\in\mathbb{N}}\bigcap_{j\geq k} A_j}(x) = \sup_{k\in\mathbb{N}} I_{\bigcap_{j\geq k} A_j} = \sup_{k\in\mathbb{N}} \inf_{j\geq k} I_{A_j} = \liminf_{j\to\infty} I_{A_j}(x)$. Using similar arguments we get the second equivalence. □

7 Vector spaces and subspaces

A vector space V as an arbitrary set together with two operations, called addition $(+)$ and scalar multiplication $(\cdot)$. Addition associates with any two elements $u, v \in V$ another element $s = u + v$ in V, called the addition (or the sum) of u and v. Scalar multiplication associates with any $u \in V$ another element $p = c \cdot u$ in V called the scalar product of u and c. If $c \in \mathbb{R}$, V is called a real vector space and if $c \in \mathbb{C}$, where $\mathbb{C}$ is the set of complex numbers, V is called a complex vector space.[4] The elements of a vector space are called vectors. We assume:

Assumption 1. *1.* $x + y = y + x$,

2. $(x + y) + z = x + (y + z)$,

3. there exist an element $\theta \in V$ *(null vector) such that for all* $v \in V$, $v + \theta = v$,

4. $c(x + y) = cx + cy$, *for* $a, c \in \mathbb{R}$ *or* $a, c \in \mathbb{C}$,

[4] For a review of fundamental properties of complex numbers see Apostol (1974).

5. $(a+c)x = ax + cx$, *for* $a, c \in \mathbb{R}$ *or* $a, c \in \mathbb{C}$,

6. $a(cv) = (ac)v$ *for* $a, c \in \mathbb{R}$ *or* $a, c \in \mathbb{C}$,

7. $0v = \theta$, $1v = v$ *for all* $v \in V$.

By $d = u - v$ we mean $u + (-1)v$ and call d the subtraction of v from u, or u minus v.

Theorem 21. *For all* $u, v, x \in V$ *and* $a, c \in \mathbb{R}$, *we have: 1.* $u + v = u + x$ *implies* $v = x$; *2.* $cv = cu$ *and* $c \neq 0$ *implies* $v = u$; *3.* $av = cv$ *and* $v \neq \theta$ *implies* $a = c$; *4.* $(a-c)v = av - cv$; *5.* $a(u-v) = au - av$; *6.* $a\theta = \theta$.

Proof. First, note that from assumption 1 - 5, $(1 + (-1))x = 0x$, and from 1 - 7 we have $0x = \theta$. Hence, $x - x = \theta$. Also, if $x - y = \theta$, by assumption 1 - 3 we have $y + x - y = y$ which implies $x = y$.

Now, 1. $u + v = u + x + (v - v) = (u + v) + x - v$ which implies that $x - v = \theta$, and consequently $x = v$; 2. $cv - cu = c(v - u) = \theta$, and if $c \neq 0$, then $v - u = \theta$; 3. $av = cv$ implies $(a - c)v = \theta$. Since, $v \neq \theta$, it must be that $a - c = 0$ and consequently, $a = c$. The proofs for items 4, 5 and 6 are left as an exercise. □

Example 8. *Let* $V = \{f : f : (0,1) \to \mathbb{R}$ *and* f *is continuous on* $(0,1)\}$. *Recall from elementary Calculus that* f *continuous on* $(0,1)$ *means that for every* $x_0 \in (0,1)$ *and for every* $\epsilon > 0$ *there exists* $\delta_{\epsilon,x} > 0$ *such that whenever* $|x - x_0| < \delta_{\epsilon,x}$ *we have* $|f(x) - f(x_0)| < \epsilon$. *Now, note that if* $f, g \in V$ *and we define* $s(x) = f(x) + g(x)$, *then by the Triangle Inequality and continuity of* f *and* g

$$|s(x) - s(x_0)| = |f(x) + g(x) - f(x_0) - g(x_0)| \leq |f(x) - f(x_0)| + |g(x) - g(x_0)| \leq \epsilon + \epsilon = 2\epsilon$$

whenever $|x - x_0| < \delta_{\epsilon,x}$. *Hence,* $s \in V$. *Also, if we define* $p(x) = cf(x)$ *for* $c \in \mathbb{R}$, *then, by continuity of* f

$$|p(x) - p(x_0)| = |cf(x) - cf(x_0)| = |c||f(x) - f(x_0)| \leq |c|\epsilon$$

whenever $|x - x_0| < \delta_{\epsilon,x}$. Hence, $p \in V$ and, if we let the null vector be the function $h(x) = 0$ for all $x \in (0, 1)$ we can conclude that V is a vector space.

Next, we provide a natural definition for addition and scalar multiplication on cartesian products of vector spaces.

Definition 13. *Let $\mathbb{X}$ and V be vector spaces. Addition and scalar multiplication are defined on the cartesian product $\mathbb{X} \times V$ as follows: $(x_1, v_1) + (x_2, v_2) = (x_1 + x_2, v_1 + v_2)$ and $a(x_1, v_1) = (ax_1, av_1)$, for any $a \in \mathbb{R}, \mathbb{C}$.*

This definition can be extended to the cartesian product $\times_{i=1}^{n} V_i$, where $n \in \mathbb{N}$ and V_i is a real vector space. In this case, $(v_1^1, \cdots, v_n^1) + (v_1^2, \cdots, v_n^2) = (v_1^1 + v_1^2, \cdots, v_n^1 + v_n^2)$ and $a(v_1, \cdots, v_n) = (av_1, \cdots, av_n)$.

Definition 14. *A nonempty subset M of a vector space V is called a subspace of V if for all $v_1, v_2 \in M$ we have $a_1 v_1 + a_2 v_2 \in M$ for any two scalars a_1 and a_2.*

Since $M \neq \emptyset$, there exists $v \in M$. Consequently, by definition $0v = \theta \in M$. It should be clear that a subspace is itself a vector space.

Example 9. *1. Let $\mathbb{X} = \mathbb{R}^2$ and define for $x, y \in \mathbb{R}^2$ the addition $s = x + y = (x_1 + y_1, x_2 + y_2)$, the scalar product $cx = (cx_1, cx_2)$ for $c \in \mathbb{R}$ and $\theta = (0, 0)$. Now, consider $M = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\}$. Clearly, $\theta \in M$. For scalars a and b and $x, y \in M$ we have $ax + by = (ax_1, ax_1) + (by_1, by_1) = (ax_1 + by_1, ax_1 + by_1) = (z, z) \in M$. Hence, M is a subspace of $\mathbb{X}$.*

2. Let $M = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\}$. Clearly, $\theta \in M$. For scalars a and b and $x, y \in M$ we have $ax + by = (ax_1, 0) + (by_1, 0) = (ax_1 + by_1, 0) = (z, 0) \in M$. Hence, M is a subspace of $\mathbb{X}$.

Definition 15. *A subset S of a vector space is said to be convex if given any $x_1, x_2 \in S$, all points written as $ax_1 + (1 - a)x_2 \in S$ when $0 < a < 1$.*

Theorem 22. *Let M and N be subspaces of the vector space $\mathbb{X}$. Then, $M \bigcap N$ is a subspace of $\mathbb{X}$.*

Proof. The null vector $\theta \in M \bigcap N$ since $\theta \in M$ and $\theta \in N$. If $x, y \in M \bigcap N$ then $x, y \in M$ and $x, y \in N$. But $x, y \in M \Rightarrow z = \alpha x + \beta y \in M$ and $x, y \in N \Rightarrow z = \alpha x + \beta y \in N$, for any scalars α, β. Therefore, $\alpha x + \beta y \in M \bigcap N$. □

Definition 16. *Let S and T be subsets of a vector space $\mathbb{X}$. $S + T$ is the set formed by all elements $s + t$ such that $s \in S$ and $t \in T$.*

Theorem 23. *Let M and N be subspaces of the vector space $\mathbb{X}$. Then, $M + N$ is a subspace of $\mathbb{X}$.*

Proof. Since M and N are subspaces, they both contain the null vector θ. Since $\theta + \theta = \theta$ we have that $\theta \in M + N$. Now, let $x, y \in M + N$. Then, there exist $m_1, m_2 \in M$ and $n_1, n_2 \in N$ such that $x = m_1 + n_1$ and $y = m_2 + n_2$. Now, let $z = ax + by$ for any two scalars a and b and note that

$$z = a(m_1 + n_1) + b(m_2 + n_2) = (am_1 + bm_2) + (an_1 + bn_2) = x_m + y_n$$

where $x_m = (am_1 + bm_2) \in M$ and $y_n = (an_1 + bn_2) \in N$. □

Definition 17. *Let $x_1, x_2, \cdots, x_n \in \mathbb{X}$ a vector space with $n \in \mathbb{N}$. A linear combination of the vectors $x_1, x_2, \cdots, x_n$ is a sum of the form $s = \sum_{i=1}^{n} a_i x_i$, where a_i are scalars.*

Definition 18. *a) Let S be a subset of the vector space $\mathbb{X}$. $[S]$ denotes the subspace generated by S. It consists of all vectors which are linear combinations of vectors in S which belong to $\mathbb{X}$. That is,*

$$[S] = \left\{ x \in \mathbb{X} : x = \sum_{i=1}^{n} a_i s_i,\ s_i \in S \text{ and } a_i \text{ scalars} \right\}.$$

b) A linear variety (or flat, affine subspace, linear manifold) V is the translation of a subspace. That is,

$$V = x_0 + M \text{ where } M \text{ is a subspace of } \mathbb{X} \text{ and } x_0 \in \mathbb{X} \text{ a vector space.}$$

c) The linear variety generated by S, say $v(S)$, is the intersection of all linear varieties in $\mathbb{X}$ that contain S.

Remark 3. *Part a) of the definition suggests that $[S]$ is a subspace. This is easy to verify since by taking $a_i = 0$, we have $x = \theta$. Also, if $x, x' \in [S]$, then for any two scalars α and β, $\alpha x + \beta x' = \alpha \left(\sum_{i=1}^n a_i s_i\right) + \beta \left(\sum_{i=1}^n a_i' s_i\right) = \sum_{i=1}^n (\alpha a_i + \beta a_i') s_i \in [S]$.*

Example 10. *1. Let $\mathbb{X} = \mathbb{R}^n$, $n \in \mathbb{N}$ and consider the vector $e_i \in \mathbb{R}^n$ where e_i has the number 1 on its i^{th} position and the number 0 in all other positions. Now define $S = \{e_i\}_{i=1}^n$. Then, $[S] = \{x \in \mathbb{X} : x = \sum_{i=1}^n e_i x_i\}$ where x_i are scalars.*
2. Consider M from Example 9 and define $\mathcal{V} = x_0 + M$ with $x_0 = (2,1)$. Then, $\mathcal{V} = \{v = (v_1, v_2) : v = (m_1 + 2, m_2 + 1)\}$ where $m = (m_1, m_2) \in M$.

7.1 Linear independence and dimension

Definition 19. *a) A vector $x \in \mathbb{X}$ a vector space is said to be linearly dependent on a set S of vectors if $x = \sum_{i=1}^n a_i s_i$, $s_i \in S$ and a_i scalars. That is, x is linearly dependent on S if $x \in [S]$.*
b) A set of vectors S is said to be linearly independent, if each $x \in S$ is not linearly dependent (independent) of the remaining vectors in S.

Theorem 24. *Let $S = \{x_1, x_2, \cdots, x_n\} \subset \mathbb{X}$. S is linearly independent if, and only if, $\sum_{i=1}^n a_i x_i = \theta$ implies $a_i = 0$ for all i.*

Proof. The theorem describes an equivalence, namely, S linearly independent is the same as $\sum_{i=1}^n a_i x_i = \theta$ implies $a_i = 0$ for all i. This requires the establishment of two implications:

1) that S linearly independent implies $\sum_{i=1}^n a_i x_i = \theta$ implies $a_i = 0$ for all i and that 2) $\sum_{i=1}^n a_i x_i = \theta$ implies $a_i = 0$ for all i implies S is linearly independent. We will establish both by proving their corresponding contrapositive statements. Recall that statement $A \Rightarrow B$ if, and only if, not $B \Rightarrow$ not A. First, suppose $\sum_{i=1}^n a_i x_i = \theta$ but for $i = r$, $a_r \neq 0$ (not B). Then, $x_r + \sum_{i=1, i\neq r}^n \frac{a_i}{a_r} x_i = \theta$, which implies that $x_r = \sum_{i=1,i\neq r}^n -\frac{a_i}{a_r} x_i$, and S is linearly dependent (not A). Second, suppose $x_r = \sum_{i=1,i\neq r}^n a_i x_i$ (not A). Then, $\sum_{i=1,i\neq r}^n a_i x_i - x_r = \theta$, and at least one scalar, $a_r = -1 \neq 0$ (not B). □

A direct consequence of this theorem is that if $S = \{x_1, x_2, \cdots, x_n\} \subset \mathbb{X}$ is a linearly independent collection, $z_1 = \sum_{i=1}^n a_i x_i$, $z_2 = \sum_{i=1}^n b_i x_i$ and $z_1 = z_2$, it must be that $a_i = b_i$ for all i.

Definition 20. *A finite set S of linearly independent vectors is said to be a basis for the space $\mathbb{X}$ if S generates $\mathbb{X}$. A vector space having finite bases is said to be finite dimensional. All other vector spaces are said to be infinite dimensional.*

Theorem 25. *Any two bases for a finite dimensional vector space contain the same number of elements.*

Proof. Suppose $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_m\}$ are bases for a vector space V with $m \geq n$. Since $\{x_1, \cdots, x_n\}$ is a basis, $y_1 = \sum_{i=1}^n a_i x_i$ for some collection of scalars $\{a_1, \cdots, a_n\}$. Since $y_1 \neq \theta$ at least one $a_i \neq 0$. Suppose, without loss of generality, that $a_1 \neq 0$. Then, $x_1 = y_1/a_1 - \sum_{i=2}^n (a_i/a_1) x_i$. Now, the collection $\{y_1, x_2, \cdots, x_n\}$ is a basis for V since for any $v \in V$ we have $v = \sum_{i=1}^n x_i b_i = x_1 b_1 + \sum_{i=2}^n x_i b_i = (y_1/a_1 - \sum_{i=2}^n (a_i/a_1) x_i) b_1 + \sum_{i=2}^n x_i b_i = y_1(b_1/a_1) + \sum_{i=2}^n (b_i - a_i b_1/a_1) x_i$. Repeating this procedure, suppose that $x_1, x_2, \cdots x_{k-1}$ have been replaced by $y_1, y_2, \cdots y_{k-1}$, such that

$$y_k = \sum_{i=1}^{k-1} a_i y_i + \sum_{i=k}^{n} \beta_i x_i.$$

Since $y_1, \cdots y_k$ are linearly independent not all $\beta_i = 0$. Without loss of generality, assume $\beta_k \neq 0$, then x_k can be written as a linear combination of $\{y_1, \cdots, y_k, x_{k+1}, \cdots, x_n\}$, which generates V. By induction on k we can replace all n of the $x_i's$ by $y_i's$ forming a generating set for V at each step. Hence, $\{y_1, \cdots, y_n\}$ generates V, and since $\{y_1, \cdots y_m\}$ is linearly independent we must have $m = n$. □

Theorem 26. *Let $\{v_1, ..., v_n\}$ be a basis for a vector space $\mathbb{X}$. Then, no set of more than n vectors in $\mathbb{X}$ is linearly independent.*

Proof. Let $\{x_1, ..., x_n, x_{n+1}\}$ be a collection of $n+1$ vectors in $\mathbb{X}$. Since $\{v_1, ..., v_n\}$ is a basis for $\mathbb{X}$, there exists α_{ik} for $i = 1, 2, ..., n$ such that $x_k = \sum_{i=1}^{n} \alpha_{ik} v_i$ for $k = 1, 2, ..., n+1$. Now, consider a linear combination of $\{x_1, ..., x_n, x_{n+1}\}$, i.e.,

$$\sum_{k=1}^{n+1} \beta_k x_k = \sum_{k=1}^{n+1} \beta_k \sum_{i=1}^{n} \alpha_{ik} v_i = \sum_{i=1}^{n} \left(\sum_{k=1}^{n+1} \beta_k \alpha_{ik} \right) v_i.$$

Linear independence of the collection $\{x_1, ..., x_n, x_{n+1}\}$ means that $\sum_{k=1}^{n+1} \beta_k x_k = \theta$ implies $\beta_k = 0$ for all k. Since the collection $\{v_i\}$ is linearly independent, then $\sum_{k=1}^{n+1} \beta_k \alpha_{ik} = 0$ for all i if $\sum_{k=1}^{n+1} \beta_k x_k = \theta$. Note that

$$\sum_{k=1}^{n+1} \beta_k \alpha_{ik} = 0 \text{ for all } i$$

is a (linear) system of n equations with $n+1$ unknowns. These systems always have non-trivial (different from θ) solutions. Hence, there exists $\beta_1, \beta_2, ..., \beta_{n+1}$ (not all zero) such that $\sum_{k=1}^{n+1} x_k \beta_k = 0$. □

Remark 4. *The number of elements in a basis for a finite dimensional space $\mathbb{X}$ is called the dimension of the space and we denote it by $dim(\mathbb{X}) \in \mathbb{N}$.*

Corollary 3. *Let $\mathbb{X}$ be a vector space such that $n = dim(\mathbb{X})$. Then, any linearly independent family of n vectors in $\mathbb{X}$ is a basis for $\mathbb{X}$.*

Proof. Let $\{b_1, b_2, ..., b_n\} \in \mathbb{X}$ be such that $\{b_1, b_2, ..., b_n\}$ is linearly independent, and let x be an arbitrary element of $\mathbb{X}$. We must show that $x = \sum_{i=1}^{n} \alpha_i b_i$ where α_i are scalars not all equal to zero. By the previous theorem, $\{b_1, b_2, ..., b_n, x\}$ forms a linearly dependent family. Therefore, $\sum_{i=1}^{n} b_i \alpha_i + \alpha_{n+1} x = \theta$ for $\alpha_1, \alpha_2, \cdots, \alpha_{n+1}$ not all zero. In particular, $\alpha_{n+1} \neq 0$, otherwise $\sum_{i=1}^{n} b_i \alpha_i = \theta$ implying $\{b_1, b_2, ..., b_n\}$ is a linearly dependent collection. Therefore, $x = \sum_{i=1}^{n} (-\frac{\alpha_i}{\alpha_{n+1}}) b_i$. □

8 Metric spaces, normed vector spaces and inner product spaces

We start by defining a metric on an arbitrary set $\mathbb{X}$. It allows us to think of proximity of the elements in a set.

Definition 21. *Let $\mathbb{X}$ be an arbitrary set. A pseudo-metric for $\mathbb{X}$ is a function $d : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ such that, for all $x, y, z \in \mathbb{X}$ it satisfies*

1. *for all $x \in \mathbb{X}$, $d(x, x) = 0$,*
2. *for all $x, y \in \mathbb{X}$, $d(x, y) = d(y, x)$,*
3. *for all $x, y, z \in \mathbb{X}$, $d(x, z) \leq d(x, y) + d(y, z)$.*

If, in addition, $d(x, y) = 0 \implies x = y$, d is called a metric on $\mathbb{X}$.

The pair $(\mathbb{X}, d)$ is called a pseudo-metric or metric space depending on the nature of d. For every $r > 0$, the set $B(x, r) = \{y \in \mathbb{X} : d(y, x) < r\}$ is called a ball centered at x and radius r. A ball is always non-empty as it contains its center. It contains all points in a space that are close to its center, where closeness is given by r.

A concept related to that of a semi-metric is that of a semi-norm on $\mathbb{X}$.

Definition 22. *Let $\mathbb{X}$ be an arbitrary vector space. A semi-norm for $\mathbb{X}$ is a function $\| \cdot \| : \mathbb{X} \to [0, \infty)$ satisfying:*

1. $\|x+y\| \leq \|x\| + \|y\|$ *for all* $x, y \in \mathbb{X}$ *(Triangle inequality),*

2. $\|\alpha x\| = |\alpha|\|x\|$ *for all* $x \in \mathbb{X}$ *and for all* $\alpha \in \mathbb{R}$.

In addition, if $\|x\| = 0 \iff x = \theta$, $\|\cdot\|$ *is called a norm for* $\mathbb{X}$.

Note that by property 2, for all $x \in \mathbb{X}$ and for $\alpha = 0$, $\|0 \cdot x\| = \|\theta\| = 0 \cdot \|x\| = 0$. Hence, $x = \theta \implies \|x\| = 0$. Thus, for $\|\cdot\|$ to be a norm it suffices to have $\|x\| = 0 \implies x = \theta$. A (semi) normed vector space will be denoted by the pair $(\mathbb{X}, \|\cdot\|)$.

A semi-normed vector space naturally becomes a semi-metric space by letting $d(x, y) = \|x - y\|$. This is easily seen by noting that $d(x, x) = \|\theta\| = 0$, $d(x, y) = \|x - y\| = \|(-1)(y - x)\| = \|y - x\| = d(y, x)$, and

$$d(x, z) = \|x - z\| = \|x - y + y - z\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z).$$

If $\|\cdot\|$ is a norm, then $d(x, y) = \|x - y\| = 0 \implies x - y = \theta \implies x = y$, which means that d is a metric. Hence, if $(\mathbb{X}, \|\cdot\|)$ is a normed vector space, $(\mathbb{X}, d)$ where $d(x, y) = \|x - y\|$ is a metric space.

The following lemma, which is a direct consequence of the triangle inequality, will be useful in subsequent proofs.

Lemma 1. *Let* $(\mathbb{X}, \|\cdot\|)$ *be a normed vector space. Then,* $\|x\| - \|y\| \leq \|x - y\|$ *for all* $x, y \in \mathbb{X}$.

Proof. $\| x \| - \| y \| = \| x - y + y \| - \| y \| \leq \| x - y \| + \| y \| - \| y \| = \| x - y \|$. □

Example 11. 1. *It is easy to verify that the absolute value of a real number is a norm on* $\mathbb{R}$ *and we have that* $(\mathbb{R}, |\cdot|)$ *is a normed vector space and* $(\mathbb{R}, d)$ *is a metric space for* $d(x, y) = |x - y|$. *In this metric space, a ball centered in* $x \in \mathbb{R}$ *with radius* $r > 0$ *is* $\{y \in \mathbb{R} : |y - x| < r\} = (x - r, x + r)$, *an open interval on* $\mathbb{R}$.

2. *The function* $\|x\|_E = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$ *is a norm on* $\mathbb{R}^n$*. To see this, note that for any* $a \in \mathbb{R}$, $\|ax\|_E = \left(\sum_{i=1}^n a^2 x_i^2\right)^{1/2} = |a|\|x\|_E$*. Also,*

$$\begin{aligned}
\|x+y\|_E^2 &= \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 + 2\sum_{i=1}^n x_i y_i \le \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 + 2\left|\sum_{i=1}^n x_i y_i\right| \\
&\le \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 + 2\left(\sum_{i=1}^n x_i^2\right)^{1/2}\left(\sum_{i=1}^n x_i^2\right)^{1/2} \\
&= \left(\left(\sum_{i=1}^n x_i^2\right)^{1/2} + \left(\sum_{i=1}^n y_i^2\right)^{1/2}\right)^2 = (\|x\|_E + \|y\|_E)^2.
\end{aligned}$$

Hence, $\|x+y\|_E \le \|x\|_E + \|y\|_E$*. Lastly,* $\|x\|_E = 0 \implies x_i = 0$ *for all* i*, establishing that* $x = \theta$*. The pair* $(\mathbb{R}^n, \|\cdot\|_E)$ *is a normed vector space, called the Euclidean space,* $\|\cdot\|_E$ *is called the Euclidean norm and* $d(x,y) = \|x-y\|_E$ *is the Euclidean metric.* $(\mathbb{R}^n, d)$ *is a metric space.*

3. *Let* $C[a,b]$ *be the collection of all real valued continuous functions defined on* $[a,b]$*. We will later define, in general terms, what is meant by a continuous function on a set* A*, but as we have discussed earlier, in this case we can rely on elementary Calculus for a notion of continuity.* $C[a,b]$ *is a vector space and we let* $\|f\| = \sup_{a\le t\le b}|f(t)|$*. Condition 1 for a norm is obviously met. For condition 2, note that*

$$\|(f+g)\| = \|f+g\| = \sup_{a\le t\le b}|f(t)+g(t)| \le \sup_{a\le t\le b}|f(t)| + \sup_{a\le t\le b}|g(t)| = \|f\| + \|g\|.$$

For condition 3, note that $\|(\alpha f)\| = \|\alpha f\| = \sup_{a\le t\le b}|\alpha f(t)| = |\alpha|\|f\|$*. The metric corresponding to this norm is* $d(f,g) = \sup_{a\le t\le b}|f(t)-g(t)|$

4. *Consider* $C[a,b]$ *and from calculus define* $\|f\| = \int_a^b |f(t)|dt$ *to be the Riemann integral of the absolute value of* f*.*[5] *The existence of* $\|f\|$ *is not in question here as* $f \in C[a,b]$ *is sufficient for the existence of* $\int_a^b |f(t)|dt$*. It can be promptly verified that the set*

[5] Later in these notes we will study the Riemann integral in detail.

of Riemann integrable functions on $[a, b]$ form a vector space and that $\|f\|$ satisfies conditions 1-3 in Definition 22. The metric corresponding to this norm is $d(f, g) = \int_a^b |f(t) - g(t)|dt$.

We now provide a definition for an *open* set associated with a metric space.

Definition 23. *Let $(\mathbb{X}, d)$ be a metric space and $U \subset \mathbb{X}$. U is called an open set if for all $x \in U$ there exists $B(x, r)$ such that $B(x, r) \subset U$.*

Definition 24. *Let $(\mathbb{X}, d)$ be a metric space and $S \subset \mathbb{X}$. $s \in S$ is said to be an interior point of S if there exists $r > 0$, such that $B(s, r) \subset S$. The set of interior points of S is denoted by $\overset{\circ}{S}$ or int(S).*

By definition, $\text{int}(S)$ is an open set and by construction $\text{int}(S) \subset S$. Also, if S is open it contains all of its interior points. Hence, if S is open $S = \text{int}(S)$.

Remark 5. *Since the interior of the empty set is empty, the $\emptyset$ is equal to its interior and consequently it is an open set. Since $\mathbb{X}$ is the universal set, for any $r > 0$ and $x \in \mathbb{X}$, $B(x, r) \subset \mathbb{X}$. Hence, $\mathbb{X}$ is open.*

Theorem 27. *In any metric space $(\mathbb{X}, d)$, $B(x, r)$ is an open set.*

Proof. We must show that for every $y \in B(x, r)$ there exists $B(y, s)$ such that $B(y, s) \subset B(x, r)$. $y \in B(x, r) \implies d(x, y) < r$, hence let $s = r - d(x, y) > 0$. Then, for any $z \in B(y, s)$ we have that $d(z, x) \leq d(z, y) + d(y, x) < s + d(y, x) = r$. Hence, $z \in B(x, r)$. $\square$

Theorem 28. *In any metric space $(\mathbb{X}, d)$, U is an open set if, only if, U is the union of open balls.*

Proof. ($\implies$) U open implies that for every $u \in U$ there exists $B(u, r_u) \subset U$ for some $r_u > 0$. Hence, $U \subset \bigcup_{u \in U} B(u, r_u) \subset U$, where the last set containment follows since $B(u, r_u) \subset U$. Hence, $U = \bigcup_{u \in U} B(u, r_u)$.

($\Longleftarrow$) Suppose $U = \bigcup_{B \in \mathcal{C}} B$ where $\mathcal{C}$ is an arbitrary collection of open balls. If $\mathcal{C}$ is empty, $U = \emptyset$ and consequently open. If $\mathcal{C} \neq \emptyset$, then $U \neq \emptyset$. Every $u \in U$ belongs to some ball B, and by Theorem 27, there exists a ball $B_u \subset B \subset U$. Hence, U is open. □

Theorem 28 can be strengthened when considering the metric space $(\mathbb{R}, d)$ where $d(x, y) = |x - y|$. To this end consider the following definition.

Definition 25. *Let U be an open subset of $\mathbb{R}$. An open finite or infinite interval $I = (a, b) = \{x \in \mathbb{R} : a < x < b\}$ is called a component interval of U if $I \subseteq U$ and if there does not exist an open interval J such that $I \subset J \subseteq U$.*

Theorem 29. *Let I denote a component interval of U. If $x \in U$, then there exists I such that $x \in I$. If $x \in I$, then $x \notin J$ where J is any other component interval of S.*

Proof. $x \in U \implies x \in I$, I an open interval for some $I \subseteq U$. There may be many such intervals, but the largest is $I_x = (a(x), b(x))$, where $a(x) = \inf\{a : (a, x) \subseteq U\}$, $b(x) = \sup\{b : (x, b) \subseteq U\}$. Note, a may be $-\infty$ and b may be $+\infty$. There is no open interval $J \ni I_x \subset J \subseteq U$ and by definition I_x is a component interval of S. If J_x is another component interval containing x, $I_x \bigcup J_x$ is an open interval $\ni I_x \bigcup J_x \subseteq U$. By definition of component interval $I_x \bigcup J_x = I_x$ and $I_x \bigcup J_x = J_x$, so $I_x = J_x$. □

Theorem 30. *Let $U \subset \mathbb{R}$ be open with $U \neq \emptyset$. Then $U = \bigcup_{n=1}^{\infty} I_n$ where $\{I_1, I_2, \dots\}$ is a collection of disjoint component intervals of U.*

Proof. If $x \in U$, then x belongs to one, and only one, component interval I_x. Note that $\bigcup_{x \in U} I_x = U$ and by the definition of component intervals and the proof of the previous theorem, the collection of component intervals is disjoint. (If x belongs to I_x and J_x, both component intervals, $I_x = J_x$). Let $\{q_1, q_2, \dots\}$ be the collection of rational numbers (countable). In each component interval, there may be infinitely many of these, but among these there is exactly one with smallest index n. Define a function F, $F(I_x) = n$ if I_x contains the rational number x_n. If $F(I_x) = F(I_y) = n$ then I_x and I_y contain x_n, and $I_x = I_y$. Thus, the collection of component intervals is countable. □

Theorem 31. *Let $(\mathbb{X}, d)$ be a metric space. Any union of open sets in $\mathbb{X}$ is open. Any finite intersections of open sets in $\mathbb{X}$ is open*

Proof. If $U = \bigcup_{i \in I} G_i$ we have that $x \in U \implies$ there exists i such that $x \in G_i$. If G_i is open, then there exists $B(x, r) \subset G_i$. But in this case, $B(x, r) \subset U$ and we conclude that U is open.

If $\mathcal{I} = \bigcap_{i=1}^{n} G_i$ and $x \in \mathcal{I}$ then $x \in G_i$ for all i. Since G_i is open, then there exists $r_i > 0$ such that $B(x, r_i) \subset G_i$. Letting $r := \min_{1 \leq i \leq n} \{r_i\}$ we have that $B(x, r) \subset G_i \subset \mathcal{I}$ for all i, and we conclude that $\mathcal{I}$ is open. □

Definition 26. *Let $(\mathbb{X}, d)$ be a metric space and $S \subset \mathbb{X}$. $x \in \mathbb{X}$ is a closure (or adherent, or contact) point of S if, for all $r > 0$, $B(x, r) \bigcap S \neq \emptyset$. $\overline{S}$ is the set of closure points of S.*

It is clear from this definition that $S \subset \overline{S}$ as for every $x \in S$ and any $r > 0$, $B(x, r) \bigcap S \neq \emptyset$.

Definition 27. *Let $(\mathbb{X}, d)$ be a metric space. $S \subset \mathbb{X}$ is closed if $S = \overline{S}$.*

Definition 28. *Let $(\mathbb{X}, d)$ be a metric space. $x \in \mathbb{X}$ is a boundary point for $S \subset \mathbb{X}$ if for all $r > 0$, $B(x, r) \bigcap S \neq \emptyset$ and $B(x, r) \bigcap S^c \neq \emptyset$. The set of boundary points for the set S is denoted by ∂S.*

Theorem 32. *Let $(\mathbb{X}, d)$ be a metric space and $S \subset \mathbb{X}$. S is open if, and only if, $\mathbb{X} - S = S^c$ is closed.*

Proof. ($\implies$) Since S is open, every $p \in S$ is such that there exists $r > 0$ such that $B(p, r) \subset S$. Thus, p cannot be a closure point of $\mathbb{X} - S$. That is, if x is a closure point of $\mathbb{X} - S$ it must be in $\mathbb{X} - S$. Thus, $\mathbb{X} - S$ is closed.

($\Longleftarrow$) If $\mathbb{X} - S$ is closed, it contains all of its closure points. That is, there is no $p \in S$ such that for any $r > 0$, $B(p,r) \bigcap (\mathbb{X} - S) \neq \emptyset$. Hence, there is always an $r > 0$ such that $B(p,r) \subset S$. So, S is open. $\square$

Definition 29. *Let $(\mathbb{X}, d)$ be a metric space. $D \subset \mathbb{X}$ is said to be dense in $\mathbb{X}$ if, and only if, $\overline{D} = \mathbb{X}$. $\mathbb{X}$ is said to be separable if, and only if, it contains a countable dense subset.*

Example 12. *From the comments following Theorem 8, $\mathbb{Q}$ is a countable set. Furthermore, from Corollary 2, for any $x \in \mathbb{R}$ and any $r > 0$ there exists $q \in \mathbb{Q}$ such that $q \in B(x,r)$. Hence, $\mathbb{Q} \bigcap B(x,r) \neq \emptyset$ and $\mathbb{Q}$ is dense in $\mathbb{R}$. Since $\mathbb{Q}$ is countable, $\mathbb{R}$ is separable.*

Even more structure can be imparted on a vector space by defining inner product spaces.

Definition 30. *A vector space $\mathbb{X}$ is called an inner-product space if for all $x, y \in \mathbb{X}$, there exists a function $\langle x, y \rangle : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, called an inner product, such that for all $x, y, z \in \mathbb{X}$ and $a \in \mathbb{R}$:*

1. $\langle x, y \rangle = \langle y, x \rangle$
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
3. $\langle ax, y \rangle = a\langle x, y \rangle$, $a \in \mathbb{R}$
4. $\langle x, x \rangle \geq 0$, *for all* x
5. $\langle x, x \rangle = 0 \iff x = \theta$, *where θ is the null vector in $\mathbb{X}$.*

Any element x of an inner-product space has a natural norm defined by $\|x\| = \langle x, x \rangle^{1/2}$. To verify that $\langle x, x \rangle^{1/2}$ is a norm, note that: a) $\langle x, x \rangle^{1/2} \geq 0$; b) for $a \in \mathbb{R}$, $\langle ax, ax \rangle^{1/2} = (a\langle x, ax \rangle)^{1/2} = (a^2\langle x, x \rangle)^{1/2} = |a|\langle x, x \rangle^{1/2}$; c) first, note that for any $a \in \mathbb{R}$ we have

$$
\begin{aligned}
0 \leq \langle ax + y, ax + y \rangle = a^2\langle x, x \rangle + 2a\langle x, y \rangle + \langle y, y \rangle \text{ and setting } a = -\tfrac{\langle x,y \rangle}{\langle x,y \rangle} \\
= -\frac{\langle x, y \rangle^2}{\langle x, x \rangle} + \langle y, y \rangle
\end{aligned}
$$

which gives $\langle x,y\rangle \le \langle x,x\rangle^{1/2}\langle y,y\rangle^{1/2}$. Now,

$$\begin{aligned}\langle x+y,x+y\rangle &= \langle x,x\rangle + \langle y,y\rangle + 2\langle x,y\rangle \le \langle x,x\rangle + \langle y,y\rangle + 2\langle x,x\rangle^{1/2}\langle y,y\rangle^{1/2}\\ &= \left(\langle x,x\rangle^{1/2} + \langle y,y\rangle^{1/2}\right)^2\end{aligned}$$

and we obtain $\langle x+y,x+y\rangle^{1/2} \le \langle x,x\rangle^{1/2} + \langle y,y\rangle^{1/2}$; d)$\langle x,x\rangle^{1/2} = 0$ implies $\langle x,x\rangle = 0$, which implies $x=\theta$. Thus, every inner product space is a normed space with this norm. Furthermore, since by letting $d(x,y) = \langle x-y,x-y\rangle^{1/2}$ we have a metric space.

Theorem 33. *(Parallelogram Law) Let $(\mathbb{X},\langle\cdot,\cdot\rangle)$ be an inner product space with norm $\|x\| = \langle x,x\rangle^{1/2}$. Then, $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$.*

Proof. $\|x+y\|^2 = \langle x+y,x+y\rangle = \langle x,x\rangle + \langle y,y\rangle + 2\langle x,y\rangle$ and $\|x-y\|^2 = \langle x-y,x-y\rangle = \langle x,x\rangle + \langle y,y\rangle - 2\langle x,y\rangle$. Hence, we obtain $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$. □

Example 13. *Let $\mathbb{X} := \mathbb{R}^n$ and $\langle x,y\rangle = \sum_{i=1}^n x_i y_i$.*

9 Topological spaces

Many of the concepts and results in the previous section can be obtained without a metric. We start by defining a topology and taking an axiomatic approach to the notion of open sets.

Definition 31. *Let $\mathbb{X}$ be an arbitrary set. A topology $\mathcal{T}$ on $\mathbb{X}$ is a collection of subsets of $\mathbb{X}$ with the following properties:*

1. $\mathbb{X}, \emptyset \in \mathcal{T}$,
2. *If $G_i \in \mathcal{T}$ where $i \in I$, I an arbitrary index set, then* $\bigcup_{i\in I} G_i \in \mathcal{T}$,
3. *If $G_i \in \mathcal{T}$ where $i = 1,\ldots,n$ for $n \in \mathbb{N}$, then* $\bigcap_{i=1}^n G_i \in \mathcal{T}$.

It is common to refer to property 2 by saying that a topology is 'closed' for arbitrary unions and to property 3 by saying that a topology is 'closed' for finite intersections. A topological space is a pair $(\mathbb{X}, \mathcal{T})$. When a topology is chosen, the sets in $\mathcal{T}$ are called *open* sets. The complements of open sets $U \in \mathcal{T}$ relative to $\mathbb{X}$, i.e., $\mathbb{X} - U$ are called *closed* sets. Since $\mathbb{X}$, $\emptyset \in \mathcal{T}$, they are open sets, but since $\mathbb{X}^c = \emptyset$ and $\mathbb{X} = \emptyset^c$, they are also closed sets.

If G_i is open, then G_i^c is closed and by DeMorgan's Law we have

$$\left(\bigcup_{i \in I} G_i \right)^c = \bigcap_{i \in I} G_i^c. \tag{3}$$

Hence, the arbitrary intersection of closed sets is a closed set. Also, since

$$\left(\bigcap_{i=1}^{n} G_i \right)^c = \bigcup_{i=1}^{n} G_i^c, \tag{4}$$

we can conclude that the finite union of closed sets is a closed set.

Example 14. *Let* $(\mathbb{X}, d)$ *be an arbitrary metric space, and let the topology* $\mathcal{T}$ *be the class of all subsets of* $\mathbb{X}$ *that satisfy Definition 23. By Remark 5 and Theorem 31, this collection satisfies properties 1, 2 and 3 in Definition 31. Note also that by Theorem 32, complements of sets in this topology are closed according to Definition 27. This topology is called the "usual" topology.*

Definition 32. *A base* for a topology $\mathcal{T}$ on $\mathbb{X}$ is any collection $\mathcal{U} \subset \mathcal{T}$ such that for every $V \in \mathcal{T}$ we have that $V = \bigcup_{U \in \mathcal{U}, U \subset V} U$.

Note that since every member V of the topology can be written as the union of elements the base, the definition requires that $x \in V \implies x \in U$ for some $U \subset V$.

Example 15. *Let* $(\mathbb{X}, d)$ *be a metric space endowed with the usual topology. Then, by Theorem 28 every open set* $U = \bigcup_{B \in \mathcal{U}} B$*. Hence, the collection* $\mathcal{U}$ *of all open balls (with all possible centers and all possible radii) in a metric space is a base for the usual topology.*

A special case of this example is the metric space $(\mathbb{R}, d)$ where $d(x, y) = |x - y|$. In this case $\mathcal{U}$ is the collection of all intervals $(x - r, x + r)$ for $x \in \mathbb{R}$ and $r > 0$.

Definition 33. *Given a topological space $(\mathbb{X}, \mathcal{T})$ and $x \in \mathbb{X}$ we say that N_x is a neighborhood of x if $x \in U \subset N_x$ for some $U \in \mathcal{T}$. A collection $\mathcal{N}$ of neighborhoods of x is a neighborhood-base at x if, and only if, for every* neighborhood N_x of x, $x \in N \subset N_x$ for some $N \in \mathcal{N}$.

Note that N_x may or may not be open. If N_x is open, we call it an open neighborhood of x.

Definition 34.

1. *Let $S \subset \mathbb{X}$. $x \in \mathbb{X}$ is a closure (adherent or contact) point of S if every neighborhood of x contains a point of S. That is, for every N_x we have $N_x \bigcap S \neq \emptyset$.*
2. *The set of all closure points of S is called the closure of S and is denoted by $\overline{S}$.*
3. *$x \in \mathbb{X}$ is a limit point of S if every neighborhood of x contains infinitely many points of S.*

The reason for part 3 of Definition 34 will become clearer later in these notes.

Theorem 34. *Let T be a set of topologies associated with $\mathbb{X}$. Then, $\bigcap_{\mathcal{T} \in T} \mathcal{T}$ is a topology.*

Proof. First, note that $\emptyset, \mathbb{X} \in \mathcal{T}$ for every topology $\mathcal{T}$. Then, we immediately have that $\emptyset, \mathbb{X} \in \bigcap_{\mathcal{T} \in T} \mathcal{T}$. Second, let $G_i \in \bigcap_{\mathcal{T} \in T} \mathcal{T}$ for every $i \in I$. Then, $G_i \in \mathcal{T}$ for every $i \in I$ and $\mathcal{T} \in T$. But since $\mathcal{T}$ is a topology $\bigcup_i G_i \in \mathcal{T}$ for every $\mathcal{T} \in T$. Hence, $\bigcup_i G_i \in \bigcap_{\mathcal{T} \in T} \mathcal{T}$. The verification for the third property of a topology follows as in the proof of the second property. □

9.1 Normed spaces and topology

We now turn to a more concrete formulation. Under a suitable choice of what constitutes an open set, metric spaces can be showed to be topological spaces. In particular, let $(\mathbb{X}, \|\cdot\|)$

be a normed space and define a metric $d_{\mathbb{X}}(x,y) = \|x-y\|$ such that we have the metric space $(\mathbb{X}, d_{\mathbb{X}})$.

Definition 35. *Let $S \subset \mathbb{X}$. $p \in S$ is said to be an interior point of S if there exists $\epsilon > 0$, such that all x satisfying $\| x - p \| < \epsilon$ are in S.*

Alternatively, if we define the $B(p,\epsilon) = \{x : \| x - p \| < \epsilon\}$, p is an interior point of S if there exists $\epsilon > 0$ such that $B(p,\epsilon) \subset S$. $B(p,\epsilon)$ is called an open-ball of radius ϵ centered at p. The set of interior points of S is denoted by $\overset{\circ}{P}$ or $int(P)$.

Definition 36. *S is an open set if $S = \overset{\circ}{S}$.*

Definition 37. *Let $(\mathbb{X}, \| \cdot \|_{\mathbb{X}})$ be a normed vector space and $S \subset \mathbb{X}$. A point x_L is said to be a limit (or cluster, or accumulation) point of S if every open ball of radius $\epsilon > 0$ centered at x_L contains (at least) a point in S distinct from x_L. That is,*

$$B(x_L, \epsilon) \cap (S - \{x_L\}) \neq \emptyset.$$

The set of all limit points of S is called the derived set of S and denoted by S^D.

Every limit point is a closure point.

Remark 6. *1. It is clear that every ball centered at x_L contains infinitely many points in S. To see this, suppose there exists $B(x_L, \epsilon)$ that contains finitely many points of S, and denote the set of these points by $\{s_1, s_2, \ldots, s_n\}$. $\{\|s_j - x_L\|\}_{j=1}^n$ is a finite collection of non-negative real numbers and we can set $m := \min_{1 \leq j \leq n} \|s_j - x_L\|$. Then, $B(x_L, m/2)$ contains no elements of S. But if this the case, x_L can't be a limit point of S. This remark justifies part 3 of Definition 34.*

2. It follows from 1 that a set with a finite number of elements cannot have a limit point.

3. Since $S \subseteq \overline{S}$, if $x \in \overline{S}$ then either $x \in S$ or $x \notin S$ but in $\overline{S}$. But that is precisely S^D. Hence, $\overline{S} = S \cup S^D$.

Example 16. *Let* $A = (0,1) \cup \{2\}$*. Note that* 1 *is a closure point and a limit point of* A*.* 2 *is a closure point, but not a limit point. It is an isolated point of* A*.*

Definitions 36 and 27 together with Remarks 31 and 5 show that metric spaces are topological spaces.

Definition 38. *Let* $(\mathbb{X}, \|\cdot\|)$ *be a normed vector space. A subset* $S \subset \mathbb{X}$ *is said to be bounded if there exists* $c \in \mathbb{X}$ *and a scalar* $r > 0$ *such that* $S \subseteq B(c, r)$*.*

Note that if S is bounded, for every $s \in S$ we have that $\|s - c\| \leq r$. By the Triangle Inequality $\|s - c\| \geq \|s\| - \|c\|$. Hence, if S is bounded

$$r \geq \|s - c\| \geq \|s\| - \|c\|, \text{ which gives } \|s\| \leq r + \|c\|.$$

Hence, if S is bounded $s \in B(\theta, r + \|c\|)$ or $S \subseteq B(\theta, r + \|c\|)$.

Example 17. *Let* $(\mathbb{X}, \|\cdot\|)$ *be a normed vector space and consider the set* $S = \{x \in \mathbb{X} : \|x\| = 1\}$*. This is called the unit ball. The set* S *is closed and bounded. The fact that it is bounded follows directly from the comment following the last definition. To verify that it is closed, we need to verify that all closure points of* S *belong to* S*. If* c *is a closure point of* S *then for any scalar* $r > 0$*,* $B(c, r) \bigcap S \neq \emptyset$*. Thus, there exists* $z \in \mathbb{X}$ *such that* $\|z - c\| \leq r$ *and* $\|z\| = 1$ *(*$z \in S$*). By Lemma 1,* $\|z\| - \|c\| \leq \|z - c\| \leq r$ *and since* $\|z\| = 1$ *we have that* $1 - \|c\| \leq r$*. Also,* $\|c\| \leq 1 + r$ *and* $1 - r \leq \|c\| \leq 1 + r$*. Since* r *can be made arbitrarily small* $\|c\| = 1$*. Hence,* $c \in S$*.*

Definition 39. *Let* $\mathbb{X}$ *be a vector space. Two norms* $\|\cdot\|_1$ *and* $\|\cdot\|_2$ *on* $\mathbb{X}$ *are said to be equivalent if there exist scalars* $a, b > 0$ *such that* $a\|x\|_1 \leq \|x\|_2 \leq b\|x\|_1$ *for all* $x \in \mathbb{X}$*.*

Remark 7. *It is clear that if a specific norm defined on* $\mathbb{X}$ *is equivalent to an arbitrary norm on* $\mathbb{X}$*, then any two norms on* $\mathbb{X}$ *are equivalent. To see this, let* $\|x\|_s$ *be a specific norm and* *suppose it is equivalent to any other norm. Then, if* $\|\cdot\|$ *and* $\|\cdot\|_1$ *are norms there exist* $a, b, c, d > 0$ *such that*

$$a\|x\|_s \leq \|x\| \leq b\|x\|_s \tag{5}$$

and

$$c\|x\|_s \leq \|x\|_1 \leq d\|x\|_s. \tag{6}$$

Since, $c > 0$ *we have from* (6) *that* $\|x\|_s \leq c^{-1}\|x\|_1 \leq c^{-1}d\|x\|_s$ *and* $b\|x\|_s \leq bc^{-1}\|x\|_1$*. Hence, from* (5) *we conclude that* $\|x\| \leq b\|x\|_s \leq bc^{-1}\|x\|_1$*. Now, since* $d > 0$ *we have from* (6) *that* $d^{-1}c\|x\|_s \leq d^{-1}\|x\|_1 \leq \|x\|_s$ *and* $ad^{-1}\|x\|_1 \leq a\|x\|_s$*. Hence, from* (5) *we conclude that* $\|x\| \geq a\|x\|_s \geq ad^{-1}\|x\|_1$*. Thus,* $\|\cdot\|$ *and* $\|\cdot\|_1$ *are equivalent.*

The following theorem shows that in finite dimensional spaces any two norms are equivalent.

Theorem 35. *Let* $\mathbb{X}$ *be a vector space such that* $dim(\mathbb{X}) = n \in \mathbb{N}$*. Any two norms defined on* $\mathbb{X}$ *are equivalent.*

Proof. Since $dim(\mathbb{X}) = n$ we can define $\{e_i\}_{i=1}^n$ to be a basis for $\mathbb{X}$. Then, for all $x \in \mathbb{X}$ there exists a collection of scalars $\{a_i\}_{i=1}^n$ such that $x = \sum_{i=1}^n a_i e_i$. Let $\|\cdot\|_1 = \mathbb{X} \to [0, \infty)$ be defined by $\|x\|_1 = \sum_{i=1}^n |a_i|$. It can be easily verified that $\|\cdot\|_1$ is a norm.

If $\|\cdot\| : \mathbb{X} \to [0, \infty)$ is any other norm, from Remark 7, it suffices to show that $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent, i.e., there exist $a, b > 0$ such that $a\|x\|_1 \leq \|x\| \leq b\|x\|_1$ for every $x \in \mathbb{X}$. Now, if $x = \theta$ (null vector on $\mathbb{X}$) the result follows trivially. Hence, assume that $x \neq \theta$ and note that $\|x\|_1 > 0$. Then, we need to show (equivalently) that

$$a \leq \frac{\|x\|}{\|x\|_1} \leq b \iff a \leq \left\|\frac{x}{\|x\|_1}\right\| \leq b \iff a \leq \|u\| \leq b$$

where $u := \frac{x}{\|x\|_1}$ and $\|u\|_1 = 1$. Note that $\|x\| = \|\sum_{i=1}^n a_i e_i\| \leq \sum_{i=1}^n |a_i|\|e_i\|$, and since the space $\mathbb{X}$ has finite dimension we can define $b := \max_{1 \leq i \leq n} \|e_i\|$ and write $\|x\| \leq b\sum_{i=1}^n |a_i| = b\|x\|_1$ or $\|u\| \leq b$.

By Lemma 1, for any $x, y \in \mathbb{X}$ we have $|\|y\| - \|x\|| \leq \|y - x\| \leq b\|y - x\|_1$. Consequently, the function $\|u\| : S \to [0, \infty)$ where $S = \{u : \|u\|_1 = 1\}$ is continuous under the norm $\| \cdot \|_1$. By the arguments in Example 17, S is closed and bounded. Since $\|u\|$ is continuous on a closed and bounded set $\|u\|$ has a minimum on S.[6] Hence, there exists $a > 0$ such that $a \leq \|u\|$, which completes the proof. □

10 Linear functions

Definition 40. *A function $f : \mathbb{X} \to \mathbb{Y}$ is linear if for all $x_1, x_2 \in \mathbb{X}$ and scalars $\alpha, \beta \in \mathbb{R}$, we have $f(\alpha x_1 + \beta x_2) = \alpha f(x_1) + \beta f(x_2)$.*

As a matter of terminology, if $\mathbb{Y} = \mathbb{R}$, then we call f a functional.

Example 18. *1. Let $\mathbb{X} = \mathbb{R}^n$ for $n \in \mathbb{N}$ and $\mathbb{Y} = \mathbb{R}$. For fixed $a \in \mathbb{R}^n$ define $f(x) = \langle a, x \rangle$ as in Definition 30. f is a linear functional. To see this, let $z = \alpha x + \beta y$ where $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$. Then,*

$$f(z) = \langle a, z \rangle = \langle a, \alpha x + \beta y \rangle = \langle a\alpha, x \rangle + \langle a\beta, y \rangle = \alpha \langle a, x \rangle + \beta \langle a, y \rangle = \alpha f(x) + \beta f(y).$$

Note that for $\mathbb{R}^n$, if we use an Euclidean norm, we have $\langle a, x \rangle = \sum_{i=1}^n a_i x_i$ where $a := (a_1, \cdots, a_n)^T$ and $x := (x_1, \cdots, x_n)^T$. In this context, we also write $\langle a, x \rangle = a^T x$.

2. Let $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$ for $n, m \in \mathbb{N}$. For fixed $a_j \in \mathbb{R}^n$ for $j = 1, \ldots, m$ define $f_j(x) = \langle a_j, x \rangle$ as in Definition 30 and $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f(x) = \begin{pmatrix} f_1(x) & f_2(x) & \cdots & f_m(x) \end{pmatrix}^T$. Let $z = \alpha x + \beta y$ where $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$. Then,

$$f(z) = \begin{pmatrix} a_1^T z & a_2^T z & \cdots & a_m^T z \end{pmatrix}^T = \alpha f(x) + \beta f(y).$$

Note that in this case we can define,

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{pmatrix} := Ax$$

[6] See Corollary 5 (Weierstrass Theorem).

where $A = \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$. *$A$ is called a matrix of dimension $m \times n$.*

Remark 8. *Recall that $0x = \theta$ for all $x \in \mathbb{X}$, where θ is the null vector in $\mathbb{X}$. Therefore, if f is linear we have $f(\theta) = f(0x) = 0f(x) = \theta$. Hence, linear functions have the property that the image of the null vector is a null vector.*

We now return to the concept of continuity at a point and give a more general definition.

Definition 41. *A function $f : (\mathbb{X}, \|\cdot\|_{\mathbb{X}}) \to (\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ is continuous at $x_0 \in \mathbb{X}$ if for all $\epsilon > 0$ there exists $\delta(x_0, \epsilon) > 0$ such that $\|f(x) - f(x_0)\|_{\mathbb{Y}} < \epsilon$ whenever $\|x - x_0\|_{\mathbb{X}} < \delta(x_0, \epsilon)$.*

The added generality of this definition, relative to that from elementary Calculus, rests on the normed vector spaces that serve as domain and co-domain for the function f and on the flexibility of the relevant norms. We emphasize that $\delta(x_0, \epsilon)$ depends on both x_0 and ϵ.

Definition 42. *Let $\{x_n\}_{n=1,2,\cdots}$ be a sequence in $(\mathbb{X}, \|\cdot\|)$. $\{x_n\}_{n=1,2,\cdots}$ is said to converge to x if $\{\|x_n - x\|\}_{n=1,2,\cdots}$ converges to zero. In this case, we write $x_n \to x$ as $n \to \infty$ or $\|x_n - x\| \to 0$ as $n \to \infty$ or $\lim_{n\to\infty} x_n = x$.*

Remark 9. *By Lemma 1, $\|x_n\| - \|x\| \leq \|x_n - x\|$ and $\|x\| - \|x_n\| \leq \|x_n - x\|$. The last inequality implies that $-(\|x_n\| - \|x\|) \leq \|x_n - x\| \Leftrightarrow \|x_n\| - \|x\| \geq -\|x_n - x\|$. Thus, $|\|x_n\| - \|x\|| \leq \|x_n - x\|$. Consequently, if $x_n \to x$, then $|\|x_n\| - \|x\|| \to 0$ or $\|x_n\| \to \|x\|$.*

The next theorem shows that if a sequence has a limit, the limit is unique.

Theorem 36. *If $x_n \to x$ and $x_n \to y$, then $x = y$.*

Proof. $\|x - y\| = \|x - x_n + x_n - y\| \leq \|x - x_n\| + \|x_n - y\| \to 0$. Hence, $\|x - y\| = 0$ and by the definition of norms $x - y = \theta$, which implies that $x = y$. □

The next theorem provides a characterization for continuity at a point.

Theorem 37. $f : (\mathbb{X}, \|\cdot\|_{\mathbb{X}}) \to (\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ *is continuous at* $x_0 \in \mathbb{X}$ *if, and only if,* $x_n \to x_0 \implies f(x_n) \to f(x_0)$.

Proof. ($\Longleftarrow$) First, recall that $x_n \to x_0$ means that for any $\delta > 0$, there exists N_δ, such that for all $n > N_\delta$ we have $\| x_n - x_0 \|_{\mathbb{X}} < \delta$. Since $\|x_n - x_0\|_{\mathbb{X}} < \delta$ implies $\|f(x_n) - f(x_0)\|_{\mathbb{Y}} < \epsilon$ we have continuity at x_0.

($\implies$) Second, suppose $x_n \to x_0$ but $f(x_n) \nrightarrow f(x_0)$. Then, there exists $\epsilon > 0$ such that for all N there exists $n > N$ such that $\|f(x_n) - f(x_0)\| \geq \epsilon$. Since $x_n \to x_0$, for all $\delta > 0$ there exists x_n such that $\| x_n - x_0 \| < \delta$ and $\| f(x_n) - f(x_0) \| \geq \epsilon$ which refutes continuity. $\square$

The previous theorem characterizes continuity at a point. If a function f is continuous at every $x \in S \subseteq \mathbb{X}$ we say that the function is continuous on S.

Theorem 38. *Let* $f : (\mathbb{X}, \|\cdot\|_{\mathbb{X}}) \to (\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ *be linear. If* f *is continuous at* x_0*, it is continuous at every* $x \in \mathbb{X}$*.*

Proof. By Theorem 37, f is continuous at $x_0 \in \mathbb{X}$ if, and only if, $x_n \to x_0 \implies f(x_n) \to f(x_0)$. So, let $x_n \to x$ for some $x \in \mathbb{X}$. Then,

$$\begin{aligned} \| f(x_n) - f(x) \|_{\mathbb{Y}} &= \|f(x_n) - f(x_0) + f(x_0) - f(x)\|_{\mathbb{Y}} \\ &= \|f(x_n - x + x_0) - f(x_0)\|_{\mathbb{Y}}, \text{ by linearity.} \end{aligned}$$

Now, since $x_n \to x$, continuity of f at x_0 guarantees that $\|f(x_n + x_0 - x) - f(x_0)\|_{\mathbb{Y}} \to 0$. $\square$

The set of all linear functions from $(\mathbb{X}, \|\cdot\|_{\mathbb{X}})$ to $(\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ will be denoted by $L(\mathbb{X}, \mathbb{Y})$. Naturally, for $f_1, f_2 \in L(\mathbb{X}, \mathbb{Y})$ we define $(f_1 + f_2)(x) = f_1(x) + f_2(x)$, $(af_1)(x) = af_1(x)$ for all $x \in \mathbb{X}$. If $x = a_1x_1 + a_2x_2$, then

$$\begin{aligned} s(x) := (f_1 + f_2)(x) &= f_1(a_1x_1 + a_2x_2) + f_2(a_1x_1 + a_2x_2) \\ &= a_1f_1(x_1) + a_2f_1(x_2) + a_1f_2(x_1) + a_2f_2(x_2) \\ &= a_1(f_1(x_1) + f_2(x_1)) + a_2(f_1(x_2) + f_2(x_2)) \\ &= a_1(f_1 + f_2)(x_1) + a_2(f_1 + f_2)(x_2) = a_1s(x_1) + a_2s(x_2), \end{aligned}$$

and

$$\begin{aligned} p(x) := (af_1)(x) &= af_1(a_1x_1 + a_2x_2) \\ &= a_1af_1(x_1) + a_2af_1(x_2) = a_1p(x_1) + a_2p(x_2). \end{aligned}$$

Consequently, the sum of two linear functions and the scalar product of a linear function are themselves linear functions. Thus, $L(\mathbb{X}, \mathbb{Y})$ is a vector space.

Definition 43. *Let $T \in L(\mathbb{X}, \mathbb{Y})$. The image of T, denoted by $im(T)$, is $im(T) := \{y \in \mathbb{Y} : y = T(x) \text{ for } x \in \mathbb{X}\}$. The null space (or kernel) of T, denoted by $null(T)$, is given by $null(T) = \{x \in \mathbb{X} : T(x) = \theta\}$.*

Theorem 39. *Let $T \in L(\mathbb{X}, \mathbb{Y})$.*

1. *$im(T)$ is a subspace of $\mathbb{Y}$.*

2. *If $\{x_1, \cdots, x_n\}$ is a basis for $\mathbb{X}$ (finite dimensional) we have that $\{T(x_1), \cdots, T(x_n)\}$ spans (generates) the $im(T)$.*

Proof. 1. We take $im(T) \neq \emptyset$. We must show that if $y_1, y_2 \in im(T)$, then for any two scalars a and b, $y = ay_1 + by_2 \in im(T)$. Since, $y_1, y_2 \in im(T)$ there exist $x_1, x_2 \in \mathbb{X}$ such that $y_1 = T(x_1)$ and $y_2 = T(x_2)$. Hence, $y = aT(x_1) + bT(x_2)$ and since T is linear, $y = T(ax_1 + bx_2)$. But since $\mathbb{X}$ is a vector space, $x = ax_1 + bx_2 \in \mathbb{X}$, hence for some $x \in \mathbb{X}$ we have $y = T(x)$. That is, $y \in im(T)$.

2. We must show that if $y \in im(T)$, then y can be written as $y = \sum_{i=1}^{n} a_i T(x_i)$ for some collection of scalars $\{a_i\}_{i=1}^{n}$. Since $\{x_1, \cdots, x_n\}$ is a basis for $\mathbb{X}$, any $x \in \mathbb{X}$ has a unique representation given by $x = \sum_{i=1}^{n} a_i x_i$. Since T is linear, $y = T(x) = \sum_{i=1}^{n} a_i T(x_i)$. □

Definition 44. *1. The rank of a set of vectors $\{x_1, \cdots, x_n\}$ is the cardinality (count) of the largest collection of independent vectors that are elements of the set.*

2. If $T \in L(\mathbb{X}, \mathbb{Y})$ and $\{x_1, \cdots, x_n\}$ is a basis for $\mathbb{X}$, the rank of T is the rank of the collection $\{T(x_1), \cdots, T(x_n)\}$.

The $rank(T)$ is the number of independent vectors in $\{T(x_1), \cdots, T(x_n)\}$.

Theorem 40. *Let $T \in L(\mathbb{X}, \mathbb{Y})$. The null space of T is a subspace of $\mathbb{X}$.*

Proof. Because T is linear, $T(\theta) = \theta$, so $null(T) \neq \emptyset$. We must show that for any $x_1, x_2 \in null(T)$ and any a and b scalars, $ax_1 + bx_2 \in null(T)$. Since $T \in L(\mathbb{X}, \mathbb{Y})$, $T(ax_1 + bx_2) = aT(x_1) + bT(x_2)$, but given that $x_1, x_2 \in null(T)$, we have $T(ax_1 + bx_2) = a\theta + b\theta = \theta$. Consequently, $ax_1 + bx_2 \in null(T)$. □

Theorem 41. *Let $T \in L(\mathbb{X}, \mathbb{Y})$ and $\mathbb{X}$ be a space with finite dimension. Then,*

$$dim(\mathbb{X}) = dim(null(T)) + dim(im(T)).$$

Proof. Let $n := dim(\mathbb{X})$, $k := dim(null(T))$ and $r := dim(im(T))$. We must show that $n = k + r$. Let $\{w_1, \cdots, w_r\}$ be a basis for $im(T)$ and $\{u_1, \cdots, u_k\}$ be a basis for $null(T)$. Since $w_i \in im(T)$, there exist $x_i \in \mathbb{X}$ such that $T(x_i) = w_i$. We will show that $\mathcal{B} = \{x_1, \cdots, x_r, u_1, \cdots, u_k\}$ is a basis for $\mathbb{X}$.

Let $x \in \mathbb{X}$, and because $\{w_1, \cdots, w_r\}$ is a basis for $im(T)$, we can write that there exists a unique collection of scalars $\{a_i\}_{i=1}^{r}$ such that $T(x) = \sum_{i=1}^{r} a_i w_i$. But since $w_i \in im(T)$, there exist x_i such that $w_i = T(x_i)$, hence $T(x) = \sum_{i=1}^{r} a_i T(x_i)$. Furthermore, since T is linear we have $T(x) = \sum_{i=1}^{r} a_i T(x_i) = T(\sum_{i=1}^{r} a_i x_i)$, which implies that $T(x - \sum_{i=1}^{r} a_i x_i) = \theta$.

Hence, $x - \sum_{i=1}^{r} a_i x_i \in null(T)$, and since $\{u_1, \cdots, u_k\}$ is a basis for $null(T)$, we can write $x = \sum_{i=1}^{r} a_i x_i + \sum_{i=1}^{k} b_i u_i$ for some unique collection of scalars $\{b_i\}_{i=1}^{k}$. Thus, x can be written as a linear combination of $\{x_1, \cdots, x_r, u_1, \cdots, u_k\}$ and it remains to be shown that this is a linearly independent collection. That is, if $\{\alpha_i\}_{i=1}^{r}$ and $\{\beta_i\}_{i=1}^{k}$ are collection of scalars such that

$$\sum_{i=1}^{r} \alpha_i x_i + \sum_{i=1}^{k} \beta_i u_i = \theta,$$

then it must be that $\alpha_i = 0$ for all $i = 1, \cdots, r$ and $\beta_i = 0$ for all $i = 1, \cdots, k$. Since T is linear,

$$T\left(\sum_{i=1}^{r} \alpha_i x_i + \sum_{i=1}^{k} \beta_i u_i\right) = \sum_{i=1}^{r} \alpha_i T(x_i) + \sum_{i=1}^{k} \beta_i T(u_i) = \theta.$$

But since $u_i \in null(T)$, we have $\sum_{i=1}^{r} \alpha_i T(x_i) = \sum_{i=1}^{r} \alpha_i w_i = \theta$. Since $\{w_1, \cdots, w_r\}$ is a basis for $im(T)$, it must be that $\alpha_i = 0$ for the last equality to hold. Hence, $\sum_{i=1}^{k} \beta_i u_i = \theta$, but since $\{u_1, \cdots, u_k\}$ be a basis for $null(T)$, it must be that $\beta_i = 0$ for the last equality to hold. Hence, $\{x_1, \cdots, x_r, u_1, \cdots, u_k\}$ is a linearly independent collection. □

Let $T \in L(\mathbb{X}, \mathbb{Y})$ and consider a certain $y \in \mathbb{Y}$. If $y = \theta$ the null vector in $\mathbb{Y}$ we have shown that there exists $x \in \mathbb{X}$, specifically $x = \theta$ (the null vector in $\mathbb{X}$) such that $T(x) = y$. If $y \neq \theta$, there are three possibilities: a) there is a unique $x \in \mathbb{X}$ such that $y = T(x)$; b) there is no $x \in \mathbb{X}$ such that $y = T(x)$, and c) there is more than one $x \in \mathbb{X}$ such that $y = T(x)$. Possibility a) holds for every $y \in \mathbb{Y}$ if, and only if, T is bijective. In this case, we say that T has an inverse, denoted by $T^{-1} : \mathbb{Y} \to \mathbb{X}$ with $x = T^{-1}(y)$.

The following theorem says that if a linear function T has an inverse T^{-1}, T^{-1} is linear.

Theorem 42. *If $T \in L(\mathbb{X}, \mathbb{Y})$ and T^{-1} exists then $T^{-1} \in L(\mathbb{Y}, \mathbb{X})$.*

Proof. We must show that for any two scalars a and b and any $y_1, y_2 \in \mathbb{Y}$, $T^{-1}(ay_1 + by_2) = aT^{-1}(y_1) + bT^{-1}(y_2)$. Since T is linear and by the definition of an inverse function

$$T^{-1}(ay_1 + by_2) = T^{-1}(aT(x_1) + bT(x_2)) = T^{-1}(T(ax_1 + bx_2)) = T^{-1}(T(x)) = x$$

for $x = ax_1 + bx_2$. But since T^{-1} exists, $T^{-1}(y_1) = x_1$ and $T^{-1}(y_2) = x_2$ and $T^{-1}(ay_1 + by_2) = x = aT^{-1}(y_1) + bT^{-1}(y_2)$. □

Theorem 43. *Let $T \in L(\mathbb{X}, \mathbb{Y})$. T is one-to-one if, and only if, $T(x) = \theta$ implies $x = \theta$.*

Proof. ($\implies$) Recall from Remark 8 that because T is linear we have that $T(\theta) = \theta$. Also, if T is one-to-one, for any other $x \in \mathbb{X}$, $T(x) \neq \theta$. Hence, $T(x) = \theta$ only when $x = \theta$.
($\Longleftarrow$) If $T(x) = \theta$ implies $x = \theta$, then the $null(T) = \{\theta\}$. For any two different $x, x' \in \mathbb{X}$, that is, $x - x' \neq \theta$, we have by linearity that $T(x) \neq T(x')$. Consequently, T is one-to-one. □

Definition 45. *Vector spaces $\mathbb{X}$ and $\mathbb{Y}$ are said to be isomorphic if there exists a linear function $T \in L(\mathbb{X}, \mathbb{Y})$ that has an inverse T^{-1}. In this case, any such T is called an isomorphism.*

Theorem 44. *Let $\mathbb{X}$ and $\mathbb{Y}$ be real vector spaces and $\{x_1, \cdots, x_n\}$ be a basis for $\mathbb{X}$.*

1. *If $\{y_1, \cdots, y_n\} \in \mathbb{Y}$, then there exists $\mathcal{T} \in L(\mathbb{X}, \mathbb{Y})$ such that $y_i = \mathcal{T}(x_i)$ for $i = 1, \cdots, n$.*

2. *If $T, f \in L(\mathbb{X}, \mathbb{Y})$ and $T(x_i) = f(x_i) = y_i$ for $i = 1, \cdots, n$, then $T(x) = f(x)$ for all $x \in \mathbb{X}$.*

Proof. 1. Consider an arbitrary (not necessarily linear) function $T : \{x_1, \cdots, x_n\} \to \mathbb{Y}$ with $T(x_i) = y_i$ for $i = 1, \cdots, n$. Since, $\{x_1, \cdots, x_n\}$ is a basis for $\mathbb{X}$, any $x \in \mathbb{X}$ can be written as $x = \sum_{i=1}^n a_i x_i$, where $a_i \neq 0$ for some i. Now, define an extension of T to all of $\mathbb{X}$ as $\mathcal{T}(x) = \sum_{i=1}^n a_i T(x_i) = \sum_{i=1}^n a_i y_i$. Note that for $w, v \in \mathbb{X}$

$$x = aw + bv = a\sum_{i=1}^n c_i x_i + b\sum_{i=1}^n d_i x_i = \sum_{i=1}^n (ac_i + bd_i)x_i.$$

Hence, $\mathcal{T}(x) = \sum_{i=1}^n (ac_i + bd_i)T(x_i) = a\sum_{i=1}^n c_i y_i + b\sum_{i=1}^n d_i y_i = a\mathcal{T}(w) + b\mathcal{T}(v)$. Hence, $\mathcal{T}$ is linear and entirely determined by $T : \{x_1, \cdots, x_n\} \to \{y_1, \cdots, y_n\}$.

2. Let $x = \sum_{i=1}^{n} a_i x_i$, and since f, T are linear functions, we have

$$\begin{aligned} T(x) &= \sum_{i=1}^{n} a_i T(x_i), \\ f(x) &= \sum_{i=1}^{n} a_i f(x_i), \end{aligned}$$

and by assumption $T(x_i) = f(x_i)$. Consequently, $T(x) = f(x)$ for all $x \in \mathbb{X}$. □

Theorem 45. *Let $\mathbb{X}$ and $\mathbb{Y}$ be real vector spaces. Then, $\mathbb{X}$ and $\mathbb{Y}$ are isomorphic if, and only if, $dim(\mathbb{X}) = dim(\mathbb{Y})$.*

Proof. ($\implies$) Let $\{x_1, \cdots, x_n\}$ be a basis for $\mathbb{X}$, then $dim(\mathbb{X}) = n$. $\mathbb{X}$ and $\mathbb{Y}$ are isomorphic if there exist $T \in L(\mathbb{X}, \mathbb{Y})$ with an inverse T^{-1}. First, we show that $\{T(x_1), \cdots, T(x_n)\}$ is a basis for $\mathbb{Y}$. That is, we show that a) $\{T(x_1), \cdots, T(x_n)\}$ is a linearly independent collection of vectors and b) any $y \in \mathbb{Y}$ can be written as $\sum_{i=1}^{n} T(x_i) a_i$. If this is the case, then $dim(\mathbb{Y}) = n$. For a), note first that $\sum_{i=1}^{n} a_i T(x_i) = \theta$ implies $T(\sum_{i=1}^{n} a_i x_i) = \theta$. Since T^{-1} exists, we have $\sum_{i=1}^{n} a_i x_i = T^{-1}(\theta) = \theta$, but given that $\{x_1, \cdots, x_n\}$ is a basis for $\mathbb{X}$ it must be that $a_i = 0$ for all i if the equality is to hold. Hence, $\sum_{i=1}^{n} a_i T(x_i) = \theta$ implies $a_i = 0$ for all i, establishing the linear independence of $\{T(x_1), \cdots, T(x_n)\}$. For b), note that for all $y \in \mathbb{Y}$ there exists one, and only one, $x \in \mathbb{X}$ such that $y = T(x)$. But $x = \sum_{i=1}^{n} a_i x_i$, hence by linearity $y = T(\sum_{i=1}^{n} a_i x_i) = \sum_{i=1}^{n} a_i T(x_i)$.

($\impliedby$) Now, suppose $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_n\}$ are bases for $\mathbb{X}$ and $\mathbb{Y}$. By Theorem 44, there exist $\mathcal{T} \in L(\mathbb{X}, \mathbb{Y})$ such that $\mathcal{T}(x_i) = y_i$. Hence, all there is to show is that $\mathcal{T}$ has an inverse $\mathcal{T}^{-1}$. Suppose $x' = \sum_{i=1}^{n} a_i x_i$ and $x'' = \sum_{i=1}^{n} b_i x_i$ are such that $\mathcal{T}(x') = \mathcal{T}(x'')$, then $\sum_{i=1}^{n} a_i \mathcal{T}(x_i) = \sum_{i=1}^{n} b_i \mathcal{T}(x_i)$ which implies that $\sum_{i=1}^{n} (a_i - b_i) \mathcal{T}(x_i) = \sum_{i=1}^{n} (a_i - b_i) y_i = \theta$. But since $\{y_1, \cdots, y_n\}$ is a basis for $\mathbb{Y}$, it must be that $a_i = b_i$ for all i, and therefore $x' = x''$. So, $\mathcal{T}$ is one-to-one, and we need only show that it is onto to conclude it has an inverse. From Theorem 39 (part 2), $\{\mathcal{T}(x_1), \cdots, \mathcal{T}(x_n)\}$ generates $\mathcal{T}(\mathbb{X})$ which is a subspace of $\mathbb{Y}$.

But $\{\mathcal{T}(x_1), \cdots, \mathcal{T}(x_n)\} = \{y_1, \cdots, y_n\}$ which is a basis for $\mathbb{Y}$. Hence, $\mathcal{T}(\mathbb{X}) = \mathbb{Y}$ and $\mathcal{T}$ is onto. □

10.1 Vector coordinates

Let $\mathbb{X}$ be a vector space of finite dimension n and $\{v_1, \cdots, v_n\}$ be a basis for $\mathbb{X}$. Then, for every $x \in \mathbb{X}$ there exists a unique collection of scalars $\{a_1, \cdots, a_n\}$ such that $x = \sum_{i=1}^n a_i v_i$. We call a_i the i^{th} coordinate of x given $\{v_1, \cdots, v_n\}$. We define the coordinate vector function as $c_v : \mathbb{X} \to \mathbb{R}^n$, such that

$$c_v(x) = c_v\left(\sum_{i=1}^n a_i v_i\right) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a.$$

Remark 10. *1. Since for any two vectors $x \neq x'$ and for a fixed basis $\{v_1, \cdots, v_n\}$, $c_v(x) \neq c_v(x')$, c_v is one-to-one.*

2. For every $a \in \mathbb{R}^n$, we have that $\sum_{i=1}^n a_i v_i$ is an element of $\mathbb{X}$ (by the definition of a vector space and the fact that $\{v_1, \cdots, v_n\}$ is a basis), thus c_v is an onto function.

3. If $x, y \in \mathbb{X}$ then $x = \sum_{i=1}^n a_i v_i$ and $y = \sum_{i=1}^n b_i v_i$. Hence, if c, d are scalars and $z = cx + dy$ we have that $z = \sum_{i=1}^n c a_i v_i + \sum_{i=1}^n d b_i v_i = \sum_{i=1}^n (c a_i + d b_i) v_i$. Thus, $c_v(z) = ca + db = c c_v(x) + d c_v(y)$. Hence, c_v is a linear function.

Let $f \in L(\mathbb{X}, \mathbb{Y})$, and $\{v_1, \cdots, v_m\}$ and $\{w_1, \cdots, w_n\}$ are basis for $\mathbb{X}$ and $\mathbb{Y}$. Then, for any $x \in \mathbb{X}$ we can write $x = \sum_{i=1}^m a_i v_i$ and because of linearity we can write $f(x) = \sum_{i=1}^m a_i f(v_i)$. Since $f(v_i) \in \mathbb{Y}$ we can write $f(v_i) = \sum_{j=1}^n b_{ji} w_j$ and $c_w(f(v_i)) = \begin{pmatrix} b_{1i} \\ \vdots \\ b_{ni} \end{pmatrix} = b_{.i}$. Since c_w is linear,

$$c_w(f(x)) = c_w\left(\sum_{i=1}^m a_i f(v_i)\right) = \sum_{i=1}^m a_i c_w(f(v_i)) = \sum_{i=1}^m a_i b_{.i} = \sum_{i=1}^m \begin{pmatrix} a_i b_{1i} \\ \vdots \\ a_i b_{ni} \end{pmatrix}.$$

Note that for fixed x (a is fixed), $f(x)$ is entirely determined by its action on v_i. This, in turn, is determined by $b_{.i}$. Hence, we can say that $f(x)$ is determined by the collection of coordinates $\{b_{.1}, \cdots, b_{.m}\}$. Thus, we can associate with $f(x)$ the following object

$$B := \begin{pmatrix} b_{.1} & \cdots & b_{.m} \end{pmatrix} = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \vdots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{pmatrix} = \begin{pmatrix} c_w(f(v_1)) & \cdots & c_w(f(v_m)) \end{pmatrix}.$$

We call this the matrix B associated with f and we write the function

$$M(f) : L(\mathbb{X}_m, \mathbb{Y}_n) \to \mathcal{M}^{n\times m},$$

where $\mathcal{M}^{n\times m}$ is the set containing matrices with n rows and m columns. Hence, we think of $M(\cdot)$ as a function that maps linear functions to a space of $n \times m$ matrices. The subscripts on $\mathbb{X}$ and $\mathbb{Y}$ in $L(\mathbb{X}_m, \mathbb{Y}_n)$ denote the dimension of the spaces.

We note that the set of matrices $\mathcal{M}^{n\times m}$ can itself be viewed as a vector space by defining addition and scalar multiplication for $Q, N \in \mathcal{M}^{n\times m}$ and $a \in \mathbb{R}$ as

1. $Q + N$ is the matrix S with element $S_{ij} = Q_{ij} + N_{ij}$
2. aQ is the matrix P with element $P_{ij} = aQ_{ij}$
3. θ, the null vector is a matrix with element $\theta_{ij} = 0$

The next theorem establishes that M is linear, one-to-one and onto. One-to-one and onto means that for any two $f, g \in L(\mathbb{X}_m, \mathbb{Y}_n)$ and $f \neq g$ we have $M(f) \neq M(g)$ and for every $M \in \mathcal{M}^{n\times m}$ there exists one, and only one, $f \in L(\mathbb{X}_m, \mathbb{Y}_n)$. Consequently, $L(\mathbb{X}_m, \mathbb{Y}_n)$ and $\mathcal{M}^{n\times m}$ are isomorphic.

Theorem 46. *Let $f, g \in L(\mathbb{X}_m, \mathbb{Y}_n)$ and $h = \alpha f + \beta g$ for scalars α and β. Then, 1. $M(h) = \alpha M(f) + \beta M(g)$; 2. M is one-to-one and onto.*

Proof. 1. Since $L(\mathbb{X}_m, \mathbb{Y}_n)$ is a vector space, $h \in L(\mathbb{X}_m, \mathbb{Y}_n)$. Let $x = \sum_{i=1}^m a_i v_i$ and observe that by Remark 10-3 (linearity of the coordinate vector function) and linearity of f and

g we have $c_w(h(x)) = \alpha c_w(f(x)) + \beta c_w(g(x)) = \alpha c_w\left(\sum_{i=1}^m a_i f(v_i)\right) + \beta c_w\left(\sum_{i=1}^m a_i g(v_i)\right) = \alpha \sum_{i=1}^m a_i c_w(f(v_i)) + \beta \sum_{i=1}^m a_i c_w(g(v_i))$. Recall that by definition

$$M(h) = \begin{pmatrix} c_w(h(v_1)) & \cdots & c_w(h(v_m)) \end{pmatrix},$$

hence

$$\begin{aligned} M(h) &= \begin{pmatrix} \alpha c_w(f(v_1)) & \cdots & \alpha c_w(f(v_m)) \end{pmatrix} + \begin{pmatrix} \beta c_w(g(v_1)) & \cdots & \beta c_w(g(v_m)) \end{pmatrix} \\ &= \alpha M(f) + \beta M(g). \end{aligned}$$

2. First, note that $M(f) = M(g)$ implies that $f(v_i) = g(v_i)$ for all $i = 1, \cdots, m$. By Theorem 44-2 this implies $f = g$. Second, $f = g$ implies $f(v_i) = g(v_i)$ for all $i = 1, \cdots, m$ since the basis are in $\mathbb{X}$. Consequently, $M(f) = M(g)$. Thus, M is one-to-one.

Now, let $Q \in \mathcal{M}^{n\times m}$ and write $Q = \begin{pmatrix} b_{.1} & \cdots & b_{.m} \end{pmatrix}$. Note that $b_{.j} = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix}$ is an array of n scalars, and if $\{w_i\}_{i=1}^n$ is a basis for $\mathbb{Y}_n$, for some $y_j \in \mathbb{Y}_n$ we have $y_j = \sum_{i=1}^n w_i b_{ij}$ since $dim(\mathbb{Y}_n) = n$. Then, if $\{v_1, \cdots, v_m\}$ is a basis for $\mathbb{X}$, by Theorem 44 there exists $\mathcal{T} \in L(\mathbb{X}_m, \mathbb{Y}_n)$ such that $\mathcal{T}(v_j) = y_j = \sum_{i=1}^n w_i b_{ij}$. Also, by Theorem 39-2, $\{\mathcal{T}(v_1), \cdots, \mathcal{T}(v_m)\}$ is a basis for the image of $\mathcal{T}$. Thus, any $y \in \mathbb{Y}_n$ can be written as $y = \sum_{j=1}^m a_j \mathcal{T}(v_j)$, and by linearity of $\mathcal{T}$, $y = \mathcal{T}\left(\sum_{j=1}^m a_j v_j\right) = \mathcal{T}(x)$, which is entirely characterized by Q. □

In the following example we will take $\mathbb{X}_n = \mathbb{R}^n$ and choose the unit-coordinate vectors as the components of the basis.

Example 19. *Let $f \in L(\mathbb{R}^m, \mathbb{R}^n)$ and choose the unit-coordinate vectors as the components of the basis for $\mathbb{R}^m$ and $\mathbb{R}^n$. That is, the basis for $\mathbb{R}^m$ is the collection $\{e_1, \ldots, e_m\}$ with*

$$e_i^T = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}$$

where 1 appears at the i^{th} position of the m-tuple. Similarly, the basis for $\mathbb{R}^n$ is the collection $\{u_1, \ldots, u_n\}$ with $u_i^T = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}$ where 1 appears at the i^{th} position of the n-

tuple. For $x \in \mathbb{R}^m$ *we have* $x = \sum_{i=1}^m x_i e_i$

$$y = f(x) = f\left(\sum_{i=1}^m x_i e_i\right) = \sum_{i=1}^m x_i f(e_i).$$

Now, $f(e_i) \in \mathbb{R}^n$*, therefore* $f(e_i) = \sum_{j=1}^n f_j(e_i) u_j = \begin{pmatrix} f_1(e_i) & f_2(e_i) & \cdots & f_n(e_i) \end{pmatrix}^T$*. Hence,*

$$\begin{aligned} f(x) &= \sum_{i=1}^m x_i \begin{pmatrix} f_1(e_i) & f_2(e_i) & \cdots & f_n(e_i) \end{pmatrix}^T \\ &= \begin{pmatrix} \begin{pmatrix} f_1(e_1) & f_1(e_2) & \cdots & f_1(e_m) \end{pmatrix} x \\ \begin{pmatrix} f_2(e_1) & f_2(e_2) & \cdots & f_2(e_m) \end{pmatrix} x \\ \vdots \\ \begin{pmatrix} f_n(e_1) & f_n(e_2) & \cdots & f_n(e_m) \end{pmatrix} x \end{pmatrix} = \begin{pmatrix} f_1(e_1) & f_1(e_2) & \cdots & f_1(e_m) \\ f_2(e_1) & f_2(e_2) & \cdots & f_2(e_m) \\ & & \vdots & \\ f_n(e_1) & f_n(e_2) & \cdots & f_n(e_m) \end{pmatrix} x \\ &:= Fx, \end{aligned}$$

where F *is an* $n \times m$ *matric with typical element* $F_{ji} := f_j(e_i)$

Example 20. *As in the previous example let* $f \in L(\mathbb{R}^m, \mathbb{R}^n)$ *and* $g \in L(\mathbb{R}^n, \mathbb{R}^p)$ *where the domain of* g *is the range of* f*. Let* $(g \circ f)(x) = g(f(x))$ *for* $x \in \mathbb{R}^m$ *and note that* $(g \circ f)(x) \in L(\mathbb{R}^m, \mathbb{R}^p)$ *(why?). Let* $\{e_1, \cdots, e_m\}$*,* $\{u_1, \cdots, u_n\}$ *and* $\{w_1, \cdots, w_p\}$ *be the unit coordinate basis for* $\mathbb{R}^m$*,* $\mathbb{R}^n$ *and* $\mathbb{R}^p$*. Then,* $f(x) = f\left(\sum_{i=1}^m x_i e_i\right) = \sum_{i=1}^m x_i f(e_i)$*,* $f(e_i) = \sum_{j=1}^n f_j(e_i) u_j$ *and* $F = [F_{ji}]_{j=1,i=1}^{n,m}$ *is the matrix of* f*. Also,* $g(y) = g\left(\sum_{j=1}^n y_j u_j\right) = \sum_{j=1}^n y_j g(u_j)$*,* $g(u_j) = \sum_{k=1}^p g_k(u_j) w_k$ *and* $G = [G_{ki}]_{k=1,i=1}^{p,n}$ *is the matrix of* g*. Now,* $(g \circ f)(x) = (g \circ f)(\sum_{i=1}^m x_i e_i) = \sum_{i=1}^m (g \circ f)(e_i) x_i$ *and*

$$\begin{aligned} (g \circ f)(e_i) = g(f(e_i)) = g\left(\sum_{j=1}^n f_j(e_i) u_j\right) &= \sum_{j=1}^n f_j(e_i) g(u_j) = \sum_{j=1}^n f_j(e_i) \sum_{k=1}^p g_k(u_j) w_k \\ &= \sum_{k=1}^p \left(\sum_{j=1}^n f_j(e_i) g_k(u_j)\right) w_k. \end{aligned}$$

Hence, the matrix of $(g \circ f)$ *is given by* $\Pi = \left[\sum_{j=1}^n f_j(e_i) g_k(u_j)\right]_{k=1,i=1}^{p,m}$*. We define this matrix to be the (Cayley) product of the matrices* G *and* F *and we write* $\Pi = GF$*. Thus, the well known formula for multiplication of matrices derives from the composition of two linear functions.*

Example 21. *1. Let* $B \subset \mathbb{R}^K$ *and* $T \in L(B, \mathbb{R}^n)$*. Let* $X \in \mathcal{M}^{n\times K}$ *be the matrix associated with* T*, such that, as in Example 19 we write* $T(b) = Xb$ *where* $b \in B$*. Let* $\{e_k\}_{k=1}^K$ *be a basis for* $\mathbb{R}^K$*, where* e_k *is the unit coordinate vector of dimension* $K \times 1$*.* $X = (X_{.1} \ \cdots \ X_{.K})$ *and* $T(e_k) = Xe_k = X_{.k}$*. The rank of* T *is the cardinality of* $\{X_{.1}, \cdots, X_{.K}\} = K$*. Recall that the null*$(T) = \{a : Xa = \theta\}$ *and by Theorem 41 we have that dim*$(null(T)) = 0$*.*
2. Note that if $y = Xb$ *and* $y = X\beta$*, we can conclude that* $X(b - \beta) = \theta$ *and this implies* $b = \beta$ *if the columns of* X *are linearly independent or, equivalently, rank*$(T) = K$*.*
3. If $Y = T(\beta) + \epsilon$*,* Y *is the element of a linear variety, not element of the image of* T*. A very interesting question is whether or not there exists an element in the image of* T *that is "closest" to* Y*. For example, is there a solution for the minimization problem*

$$\min_{\beta} \|Y - T(\beta)\|_E?$$

10.2 Bounded linear functions

We start by establishing that linear functions defined on finite dimensional spaces are always continuous.

Theorem 47. *Let* $(\mathbb{X}, \|\cdot\|_{\mathbb{X}})$ *and* $(\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ *be normed vector spaces with dim*$(\mathbb{X}) = n \in \mathbb{N}$*.* $f \in L(\mathbb{X}, \mathbb{Y})$ *implies that* f *is continuous on* $\mathbb{X}$*.*

Proof. Since $dim(\mathbb{X}) = n \in \mathbb{N}$, there exists a basis $\{b_i\}_{i=1}^n$ and a collection of scalars $\{s_i\}_{i=1}^n$ such that any $x \in \mathbb{X}$ can be written as $x = \sum_{i=1}^n s_i b_i$. Linearity of f implies $f(x) = \sum_{i=1}^n s_i f(b_i)$ and by the properties of norms,

$$\|f(x)\|_{\mathbb{Y}} \leq \sum_{i=1}^{n} |s_i| \|f(b_i)\|_{\mathbb{Y}}.$$

The set of real numbers $\{\|f(b_i)\|_{\mathbb{Y}}\}_{i=1}^n$ is finite and we define $M := \max_{1\leq i\leq n} \|f(b_i)\|_{\mathbb{Y}}$. Hence,

$$\|f(x)\|_{\mathbb{Y}} \leq M \sum_{i=1}^{n} |s_i|.$$